DISSERTATION

HERBICIDE RESISTANCE IN KOCHIA AND WHEAT: THE LOSS AND GAIN OF WEED CONTROL TOOLS FOR WHEAT PRODUCTION

Submitted by

Jacob S. Montgomery

Department of Agricultural Biology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Fall 2024

Doctoral Committee:

Advisor: Todd Gaines

Franck Dayan
Stephen Pearce
Richard Mason

ABSTRACT

HERBICIDE RESISTANCE IN KOCHIA AND WHEAT: THE LOSS AND GAIN OF WEED CONTROL TOOLS FOR WHEAT PRODUCTION

Weeds are one of the main causes of yield loss across agricultural systems worldwide. Currently, weed control in crop production systems relies heavily on the use of chemical herbicides. While these herbicides are largely effective and very efficient, they are prone to herbicide resistance evolution in weeds. Herbicide resistance evolution is the most pressing issue facing weed scientists, currently. As researchers scramble to develop new weed control technologies, herbicide resistance traits enable the use of herbicides in new cropping systems. While this is not the long-term answer to issues facing weed managers, herbicide resistance traits offer a short-term solution to systems that may be desperate for answers.

Kochia is a tumble weed that was introduced to North America in the 1800's. Since its introduction, kochia has invaded many settings of the American West. Its abiotic stress tolerance, tumble seed dispersion mechanism, and ability to outcross allow it to invade new areas and maintain genetic diversity for selection to act on. In a study reported here, I identify the genetic basis of resistance to the herbicide dicamba in a population of kochia collected from Colorado. I use linkage mapping to identify a region of the genome associated with resistance. Within this locus, I find a transposable element insertion within an exon of an *AUXIN/INDOLE-3-ACETIC ACID* gene. This insertion causes differential splicing that changes the amino acid sequence near the degron domain in resistant plants. I apply dicamba to Arabidopsis plants expressing wildtype and mutant alleles of this gene to demonstrate that this mutation is sufficient in causing dicamba resistance. Protein modeling suggests that while several amino acids are affected, a specific glycine-to-threonine substitution is likely the most important in causing resistance. Finally, the mutant allele of this gene segregates with reduced plant height and biomass, suggesting this resistance mechanism has a pleotropic effect of a fitness cost. This study demonstrates the diverse ways that adaptive alleles can be generated in weedy species and describes the genetic and physiological basis of dicamba resistance in kochia. These findings may be used to design new auxinic herbicides that are not affected by this resistance mechanism.

With the loss of herbicide efficacy in the face of herbicide resistance evolution, as described above, wheat producers could benefit from the availability of more weed control options. In another study presented here, I utilize gene editing and traditional mutagenesis approaches to generate wheat plants that are deficient in sulfolipids. In rice, eliminating these sulfolipids clearly results in resistance to the herbicide oxyfluorfen. Liquid chromatography coupled with mass spectrometry (LC/MS) experiments confirmed that the edits I made to the gene *UTP-GLUCOSE-1-PHOSPHATE URIDYLYLTRANSFERASE 3* indeed resulted in drastic reduction in sulfolipid content. Dose response experiments showed that mutant lines gained moderate resistance to oxyfluorfen, but not lactofen, a herbicide from the same chemical family. Finally, LC/MS experiments confirmed that mutant wheat plants accumulated less porphyrin following oxyfluorfen application. The wheat lines developed in this study provide germplasm for trait introgression or a framework to make similar edits in locally adapted varieties to reproduce this herbicide resistance trait. If stewarded correctly, the trait will extend the effective lifespan of currently used herbicides in wheat and improve weed control.

ACKNOWLEDGMENTS

First, I would like to thank my advisor, Dr. Todd Gaines for believing in me and bringing me into his lab, despite already having a very busy lab. You always made time to meet and discuss questions or issues I was facing and always kept a good spirit, even in the face of disappointing results. Your broad knowledge and network pushed my projects to levels I never could have achieved on my own and gave me valuable experiences that will serve me throughout my career. I would also like to thank Dr. Franck Dayan for being a valuable mentor and friend during my degree. You went beyond a scientific advisor and really made me feel supported as a person at Colorado State. Finally, thank you to Drs. Esten Mason and Stephen Pearce for serving on my committee and for helpful feedback during my degree. I didn’t know the first thing about wheat production or gene editing when I started, and you have given me perspectives that help me see crop production and breeding more completely.

Next, I would like to thank my friends and lab mates that I have worked with over the years at the weed lab. Moving away from my family and friends was difficult, but you all provided me with a community of friends that I could rely on. You have helped me remember that life is bigger than your latest experimental result or article submission decision. I have also learned as much or more from You as I did from any class or meeting with my committee. I wish you all success and am inspired by your passion and hard work.

Thank you to my family. I know that moving to Colorado has caused me to miss things in your lives, but you have never been anything but supportive for me. You inspire me to be the best version of myself. Your love for me reminds me to love others and cultivate a climate of respect in the world.

Finally, and most importantly, thank you to my soon to be wife, Catherine. You have supported me and believed in me, even at times when I did not. You moved across the country with me and have made a home out of wherever we are. You have always taken an interest in my work but remind me that life is bigger than what you work on during your career. I love you.

TABLE OF CONTENTS

INTRODUCTION

Agricultural weeds are characterized by their adverse effects on crop production. Each year, producers spend billions of dollars to control plant pests (Pimentel et al., 2000; Gianessi & Nathan, 2007). For several decades, weed control in agricultural systems has largely relied on chemical control measures that use small molecule herbicides. These herbicides are mostly synthetic, but some are derived from natural sources (Schwartz et al., 2004). Current herbicides generally act as drugs to metabolic enzymes and inhibit core biosynthesis pathways. Blocking these important biosynthesis pathways can result in starvation for important substrates or the production of harmful by-products that cause cellular damage. Some exceptions to this generality include the dinitroaniline herbicides, which block the polymerization of tubulin proteins (Morejohn et al., 1987), and auxinic herbicides, which mimic the endogenous plant hormone auxin (Song, 2014). While opponents of herbicides cite their negative effects on the environment and non-target species (Hayes et al., 2010), the efficacy and efficiency of modern herbicides have facilitated higher agricultural outputs with reduced tillage (Aktar et al., 2009).

The discovery and use of chemical herbicides has been one of the most impactful developments in modern agriculture. However, their widespread adoption makes herbicide application a massively parallel selection experiment. In the face of such extreme selection pressure, to quote Jeff Goldbloom in *Jurassic Park*, “Life, uh, finds a way”. Thus, in the years following the introduction of any major herbicide, there have been reports of weed populations that had evolved herbicide resistance. The study of these herbicide resistant populations has led to the discovery of diverse resistance mechanisms that range from the predictable, to the unimaginable. For example, single nucleotide substitutions in conserved domains of herbicide target site genes were amongst the earliest discovered resistance mechanisms (Tranel & Wright, 2002). In contrast to this relatively straight-forward mechanism of resistance, complex genomic rearrangements have also been shown to lead to herbicide resistance. For example, translocation of herbicide target site genes followed by gene amplification has led to the increase in expression of these genes in some species (Koo et al., 2018; Zhang et al., 2023). The genetic nature of herbicide resistance has implications towards the frequency of new herbicide resistance alleles appearing *de novo*. For example, Kreiner et al. (2022b) use selective sweep, ancestral recombination graph, and effective population size estimation approaches to show that single nucleotide polymorphisms and slippage in microsatellite regions that lead to herbicide resistance are expected to evolve de novo in *Amaranthus tuberculatus* every year. In contrast, sequence of *5-enolpyruvylshikimate-3- phosphate synthase* into extrachromosomal DNA in *Amaranthus palmeri* is conserved across all samples tested (Molin et al., 2020). This suggests a single origin of this relatively rare resistance allele that has proliferated across the world. Studies elucidating herbicide resistance mechanisms provide retrospective understanding of how herbicide resistance evolution plays out from a molecular to ecosystem view. They can also inform the design of new herbicides that are not affected by currently widespread resistance mechanisms (Barker et al., 2023).

As shown by Kreiner et al. (2022b) weeds are incredible sources of genetic variation. Because they grow in agricultural settings alongside crops, environmental selection pressures select for adaptive alleles within this great pool of genetic diversity. In fact, in another study, Kreiner et al. (2022a) use paired natural and agricultural sampling of *Amaranthus tuberculatus* to identify candidate adaptive alleles. This study lays a possible framework for other studies to identify adaptive alleles in other weed species as well. Functional studies can advance the hypotheses generated by comparative studies like the one reported by Kreiner et al. (2022a) to confirm the role, if any, of these candidate alleles and their potential value to crop breeders.

A reference genome assembly is a requisite for comparative studies like the one mentioned above. Until recently, this genomic resource was only available for a few weed species. The International Weed Genomics Consortium is a collaborative group of public and private researchers that have come together to develop community tools for weed genomics (Montgomery et al., 2024). Armed with these new tools, weed scientists can now address a range of new fields of research.

The present dissertation is structured into three chapters. Chapter 1 describes the formation and outputs of the International Weed Genomics Consortium. The reference genome assemblies and accompanying annotations will advance the field of weed science by informing the development of new weed control methods. Additionally, these genome assemblies will allow for comparative genomic studies to identify adaptive alleles in weeds that may be used in crop improvement, especially alleles associated with abiotic stress tolerance. Chapter 2 presents a research study that identifies the genetic basis of dicamba resistance, a specific type of abiotic stress, in the tumbleweed kochia. This chapter outlines the loss of a management tool for agricultural producers. This study uses linkage mapping to identify a locus containing a gene known to be involved in dicamba activity. A transposable element insertion within this gene causes a change in splicing in resistant plants. Arabidopsis expressing the variant allele of this gene were resistant to dicamba, and protein modeling suggests a specific amino acid substitution of a glycine to a threonine is the main driver of resistance. The final chapter discusses the development of a new herbicide resistance trait in wheat. This chapter describes the gain of a new management tool for wheat producers. Gene editing and traditional mutagenesis approaches

were used to eliminate the function of a gene involved in sulfolipid production. Sulfolipid content was nearly completely reduced in both mutant lines. Mutant lines also showed moderate levels of resistance to the herbicide oxyfluorfen. Finally, the mutant lines accumulated less of a metabolite known to accumulate as a consequence of oxyfluorfen action. Taken together, this dissertation illustrates the value of genomic tools for the study of weeds and crops alike. It provides mechanistic understanding of a new case of herbicide resistance as well as the process of developing a new herbicide resistance trait to offer wheat producers more weed control options.

References

Aktar, M.W., Sengupta, D. & Chowdhury, A. (2009) Impact of pesticides use in agriculture: their benefits and hazards. *Interdiscip Toxicol*, **2**, 1-12.

Barker, A.L., Pawlak, J., Duke, S.O., Beffa, R., Tranel, P.J., Wuerffel, J. et al. (2023) Discovery, mode of action, resistance mechanisms, and plan of action for sustainable use of Group 14 herbicides. *Weed Science*, **71**, 173-188.

Gianessi, L.P. & Nathan, P.R. (2007) The value of herbicides in U.S. crop production. *Weed Technology*, **21**, 559-566.

Hayes, T.B., Khoury, V., Narayan, A., Nazir, M., Park, A., Brown, T. et al. (2010) Atrazine induces complete feminization and chemical castration in male African clawed frogs (*Xenopus laevis*). *Proceedings of the National Academy of Sciences*, **107**, 4612-4617.

Koo, D.-H., Molin, W.T., Saski, C.A., Jiang, J., Putta, K., Jugulam, M. et al. (2018) Extrachromosomal circular DNA-based amplification and transmission of herbicide resistance in crop weed *Amaranthus palmeri*. *Proceedings of the National Academy of Sciences*, **115**, 3332-3337.

Kreiner, J.M., Latorre, S.M., Burbano, H.A., Stinchcombe, J.R., Otto, S.P., Weigel, D. et al. (2022a) Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science*, **378**, 1079-1085.

Kreiner, J.M., Sandler, G., Stern, A.J., Tranel, P.J., Weigel, D., Stinchcombe, J.R. et al. (2022b) Repeated origins, widespread gene flow, and allelic interactions of target-site herbicide resistance mutations. *eLife*, **11**, e70242.

Molin, W.T., Patterson, E.L. & Saski, C.A. (2020) Homogeneity among glyphosate-resistant Amaranthus palmeri in geographically distant locations. *PLOS ONE*, **15**, e0233813.

Montgomery, J., Morran, S., MacGregor, D.R., McElroy, J.S., Neve, P., Neto, C. et al. (2024) Current status of community resources and priorities for weed genomics research. *Genome Biology*, **25**, 139.

Morejohn, L.C., Bureau, T.E., Molè-Bajer, J., Bajer, A.S. & Fosket, D.E. (1987) Oryzalin, a dinitroaniline herbicide, binds to plant tubulin and inhibits microtubule polymerization in vitro. *Planta*, **172**, 252-64.

Pimentel, D., Lach, L., Zuniga, R. & Morrison, D. (2000) Environmental and economic costs of nonindigenous species in the United States. *BioScience*, **50**, 53-65.

Schwartz, D., Berger, S., Heinzelmann, E., Muschko, K., Welzel, K. & Wohlleben, W. (2004) Biosynthetic gene cluster of the herbicide phosphinothricin tripeptide from Streptomyces viridochromogenes Tü494. *Appl Environ Microbiol*, **70**, 7093-102.

Song, Y. (2014) Insight into the mode of action of 2,4-dichlorophenoxyacetic acid (2,4-D) as an herbicide. *Journal of Integrative Plant Biology*, **56**, 106-113.

Tranel, P.J. & Wright, T.R. (2002) Resistance of weeds to ALS-inhibiting herbicides: what have we learned? *Weed Science*, **50**, 700-712.

Zhang, C., Johnson, N., Hall, N., Tian, X., Yu, Q. & Patterson, E. (2023) Subtelomeric 5-enolpyruvylshikimate-3-phosphate synthase (*EPSPS*) copy number variation confers glyphosate resistance in *Eleusine indica*. *Nature Communications*, **14**, 4865.

CHAPTER I – HOW THE INTERNATIONAL WEED GENOMICS CONSORTIUM IS HELPING ADVANCE THE FIELD OF WEED SCIENCE INTO THE ERA OF GENOMICS

Summary

The development and use of genomic resources has accelerated the advancement of agricultural production over the past several decades. Until recently, development of these resources for weed species lagged compared to other fields. Understanding the sequence and order of weed genomes unlocks the potential for research approaches to identify the genetic basis of important traits like herbicide resistance and abiotic stress tolerance. Knowledge around such traits can inform the development of more productive crops and further our understanding of adaptation mechanisms. Comparison of reference genome assemblies for many weed species also promises to improve weed management through characterization of new herbicide target sites and development of next generation weed control methods. Here, we briefly discuss the history of genomics in weed science and discuss the formation and current outcomes of the International Weed Genomics Consortium, a collaborative group of researchers working to make genomic resources for weed species accessible and usable to the research community. We finish with a forward-looking discussion of the direction of genomics in weed science and possible uses of the resources that have recently been developed.

Genomics in Weed Science

The era of plant genomics began with the publication of the assembled genome sequence for *Arabidopsis thaliana* in 2000 (The Arabidopsis Genome Initiative, 2000). At the time, it was just the third multi-cellular eukaryotic organism to have its genome assembled (after the nematode, *C. elegans*, and fruit fly, *D. melanogaster*). As DNA sequencing technology has improved and become cheaper, chromosome-level reference genome assemblies have been reported for non-model species and species with complex genomes. Generation of a complete genome assembly has gone from a massive undertaking, requiring a dozen research labs and millions of dollars in funding to a process that can be accomplished by a single research group with access to a sequencing core and less than $100,000 USD. Figure 2-1 illustrates the reduction in sequencing cost over the last 20 years. In the early 2000's, sequencing a Palmer amaranth genome would have cost more than a home in London. Now, you can obtain the sequence data for the entire genome for less than the cost of a hamburger. It is important to note that assembling a new genome requires sequencing from more expensive technology and genotype-by-sequencing approaches require sequencing that covers the genome several times over; however, Figure 2-1 helps illustrate the rapid increase in accessibility of sequence data for individual research labs.

While *Arabidopsis thaliana* could be considered a weedy plant by some definitions, it is not a major agricultural pest. As time went on, crops that are close relatives to weeds had their genomes assembled and sometimes included sequencing of weed species, e.g. *Setaria viridis* assembled by Bennetzen et al. (2012). These early genome assembly projects were generally intended to help detect alleles for crop improvement instead of weed control and resulted in varying levels of contiguity. For example, sequence from *Setaria viridis* was used to improve *Setaria italica*, a cultivated grain and model for studying photosynthesis and plant biomass production. Reference transcriptome assemblies, like the one reported by Riggins et al. (2009), allowed for the first untargeted investigations towards the genetic basis of weedy traits and served as a sort of intermediate step between individual gene sequencing and genome-wide analyses in weeds. Laforest et al. (2020) were the first to report a chromosome-level, fully *de* *novo* assembly of an agricultural weed, though others had previously used genomes of closely related species to help scaffold other weed genomes.

As time went on, research groups produced genomic resources for weed species, as needed, to address questions around herbicide resistance evolution, adaptation, and crop improvement. Kreiner et al. (2022) used the waterhemp genome to identify alleles that allowed it to adapt to management and spread across North America. Onkokesung et al. (2022) developed a lateral flow test to detect herbicide resistance in blackgrass based off knowledge gained from assembling its genome. Gaines et al. (2021) used the Palmer amaranth genome to track the origins of weed populations that had been introduced to South America. This research highlighted the need for the development of assays for noxious weed detection in imported seed lots. In fact, Brusa et al. (2021) developed an assay to rapidly detect the very small seeds of Palmer amaranth that are otherwise indistinguishable from seeds of other amaranths in seed mixes. This assay and many similar assays to detect alleles that cause herbicide resistance in weeds are offered by diagnostic clinics that can certify seed lots and help producers make in-season management decisions.

While genome assemblies have allowed for investigations into aspects of weediness for individual species or genus, sequence information from diverse weed species promises to uncover genetic themes of weediness in a broader sense. Similar approaches have already been used in mammals to understand evolutionary constraint and predict genetic sites of human disease (Vignieri, 2023). Insights gleaned from these large comparative projects can identify new conserved herbicide targets and inform development of next generation weed control methods.

IWGC Formation and Goals

The first discussions of an International Weed Genomics Consortium (IWGC) started as far back as 2016 at the International Weed Science Congress meeting in Prague. Discussions at weed science meetings over the next years resulted in a review paper highlighting the need for and potential of weed genome assemblies; the first official record of the IWGC (Ravet et al., 2018). Authors noticed a common need for these resources and suggested a mutually beneficial approach of communal resource generation instead of generation of private, redundant resources by individual research groups. In an unprecedented collaboration, founding members that are generally competitors in the field of crop protection (including Syngenta, Corteva Agriscience, Bayer Crop Science, and BASF) and an affiliate member (the Herbicide Resistance Action Committee) agreed to fund and produce the data necessary to assemble genomes for at least 10 weed species. The $1.5 million USD investment from the founding and affiliate members was matched by the US-based Foundation for Food and Agriculture Research to total $3 million USD to be used over three years. The goals of this consortium included generation of annotated genome assemblies for weed species, as well as to ensure accessibility and usability of these assemblies by the research community. With funding secured and legal agreements in place, the IWGC officially launched in 2021 with commencement of work on sequencing projects and its first conference in Kansas City, MO, USA.

The general inputs and outputs of the IWGC are outlined by Figure 2-2 and described in depth by Montgomery et al. (2024). At launch, executive board members from around the world voted on weed species to include in sequencing projects. This vote considered global prevalence, economic impact, and then-current availability of genomic resources for each weed species. Seeds of selected species were solicited from collaborators across the globe in accordance with the Nagoya protocol, which ensures fair and equitable sharing of benefits arising from accessing germplasm from other countries. Generally, an accession of each species that is sensitive to all herbicides was selected and grown at Colorado State University. Plant tissue was shared with several labs including Iowa State University (genome size estimation), Kansas State University (karyotyping), and Corteva Agriscience (various sequencing approaches) to produce data needed for genome assembly. This data was processed and analyzed by members of Corteva Agriscience and Michigan State University to produce genome assemblies with annotations of gene coding regions and the predicted function of their protein products. The IWGC hosts this data on an online database, called WeedPedia, and provides trainings to ensure the genomic resources are accessible and usable by the research community. The IWGC sees value in making seeds of sequenced accessions available to the public. To allow standardization of experiments across labs, the IWGC is submitting seed samples to the United State Department of Agriculture Germplasm Resources Information Network (USDA GRIN). Researchers will be able to order seed from this repository to use in their own lab.

A stipulation of the founding member agreement was that access to genome assemblies stayed private to the founding members for one year following completion. A summary of projected availability dates is provided by Montgomery et al. (2024). Teams within the IWGC are working to write manuscripts reporting each genome assembly and aim to publish these reports before the data availability date. Each manuscript will provide instructions to access the genome assembly and all raw data used for assembly or analysis. Researchers who use these tools should cite the paper reporting the genome assembly for their species of interest.

Current Outputs of the IWGC

To date, the IWGC has produced 31 genome assemblies for 19 different weed species. Species with multiple assemblies are the product of assembling the two haplotypes of outcrossing, highly heterozygous species, not from assembling the genomes of multiple plants. This data is hosted on WeedPedia, an online database developed by the IWGC and Keygene, along with genome assemblies that have been produced by research groups independent of the IWGC. In total, WeedPedia hosts 62 genome assemblies from 44 species. Figure 2-3 shows the relatedness of each of these species based off *ALS* gene sequence. The species are roughly split between monocot and dicot species and represent 10 different plant families. Gene coding regions have been annotated in all these assemblies with special denotation of herbicide target site genes. WeedPedia hosts tools for analysis such as blast and a genome browser, but data is also available for download for offline analyses. To gain access to WeedPedia, researchers can request a free account at the IWGC's website (www.weedgenomics.org). This account will grant access to all publicly available genomes, and newly completed genomes by the IWGC will be unlocked as the data availability date is reached.

To help address knowledge gaps around genomic analyses, the IWGC has hosted many trainings and workshops aimed at a variety of different skill levels. Some of these trainings have been recorded and are available at the IWGC website. The IWGC has hosted workshops in the United States, Brazil, Canada, Turkey, Australia, and England to over 300 attendees. The IWGC also hosts a monthly webinar to spotlight new advancements in the field of weed genomics. News and events are regularly posted on the IWGC website.

What's Next for Weed Genomics and the IWGC?

The first phase of the IWGC will finish near the end of 2024. Genomes will continue to be hosted on WeedPedia for the foreseeable future, but no new genome assembly projects are planned. Moving forward, the consortium looks to be more focused on curating genome annotations to improve gene model prediction and ensure complex gene families are correctly documented in the current assemblies. Trainings on common genomics methods will also be paramount as new users start to take advantage of the assemblies produced in phase one.

Now armed with genome assemblies for many of the most important agricultural weeds, researchers must prioritize their efforts to deliver on the potential of these new tools. One of the first application of these assemblies will be to understand non-target site herbicide resistance mechanisms. While much is known about target site mutations leading to herbicide resistance, the genetic basis has been determined for only a few cases of metabolic herbicide resistance. Genetic mapping approaches like linkage or association mapping will allow researchers to identify parts of the genome that cause resistance. Matched with gene expression and physiology data, researchers have been able to identify specific genes that cause metabolic herbicide resistance in weeds (Rigon et al., 2024). Identifying these metabolism genes will allow for estimation of cross resistance patterns, as shown by Bobadilla and Tranel (2024) and guide the design and selection of new herbicide compounds that will not interact with known metabolism enzymes. Similarly, structural modeling of potential herbicide target sites from all sequenced weed species will allow for *in silico* docking of new herbicides and prediction of activity spectrums, at least at the molecular level.

Thinking beyond traditional herbicide development, the resources generated by the IWGC will be vital in the development of next generation control technologies. For example, the pharmaceutical industry has found success in the development of drugs that modulate protein-

protein interactions (Lu et al., 2020), similar to herbicides like pendimethalin that disrupt tubulin polymerization. The introduction of open-source software like Alphafold3 (Abramson et al., 2024) allow us to computationally predict the structure of protein complexes with only the amino acid sequence of the proteins thought to be interacting. Understanding the interaction between proteins and other proteins/ligands then allows us to estimate molecules' abilities to disrupt important interactions. Understanding protein interactions will also help the development of herbicidal proteolysis-targeting chimera (PROTAC) molecules (Leon and Bassham, 2024). PROTACs use a bait molecule that interacts with a target protein. This bait molecule is linked to a ligand for E3 ligase proteins that are naturally produced by cells. When E3 ligase interacts with a target protein that is drawn in by the bait ligand, it is ubiquitinated and labeled for degradation, offering a new strategy for reducing protein activity/concentration. Gene sequences from many weed species can also be used to design very specific gene silencing techniques, a topic recently reviewed by Zabala-Pardo et al. (2022). Gene silencing is already being employed to control corn rootworm through the expression of a double stranded RNA molecule that targets *Snf7*, a gene necessary for transport and sorting of transmembrane proteins (Bolognesi et al., 2012). In weeds, delivery will more likely be through a spray solution that delivers RNA or DNA directly to the leaves of growing weeds. This delivery method has proved effective for a double stranded RNA product that was recently approved to control Colorado potato beetle (Rodrigues et al., 2021). While these RNAi formulations can target genes involved in core metabolic pathways, as most conventional herbicides do, they can also be designed to inhibit specific processes like floral development or seed maturation, allowing for weed management even after weed infestations are well established.

Thinking beyond applications towards the development of new control technologies, the availability of these genomic resources will facilitate studies that aim to understand the basis of desirable traits like abiotic stress tolerance. Lessons learned from identifying the genetic and physiological basis of these traits can inform crop breeders or biotechnologists in developing crops that are more resilient to changes in climatic conditions. Currently, adaptive alleles for crop improvement have been mined from diverse wild relatives of crops. In the case of triticale, biotic and abiotic stress tolerance alleles have been identified and are largely contributed by the rye progenitor (Golebiowska-Paluch and Dyda, 2023). Where variability in stress tolerance exists, genetic mapping approaches will be valuable tools in identifying genes/alleles controlling these traits. Once identified, expressing these genes or modulating expression of their endogenous ortholog in crops may lead to crop improvement, especially in cases where there is little divergence between the weedy species and crop being improved. One proven example is the development of salt tolerant alfalfa that expresses a vacular sodium transporter gene originally from the weed species *Salsola soda* (Li et al., 2010).

The availability of genomic sequence for agronomically important weed species will allow for new research to understand weed ecology and evolution. It will also help increase the effectiveness of traditional herbicide discovery pipelines through *in silico* prediction of molecule performance and specificity. These resources may also inform the development of technologies that challenge the paradigm that enzyme-inhibiting small molecules are the cornerstone of weed management. The IWGC has doubled the amount of weed species with reference genomes over the past three years and will work to improve these resources and deliver training to researchers that will help propel weed science into the era of genomics and develop tools for the sustainable management of weeds.

Figures

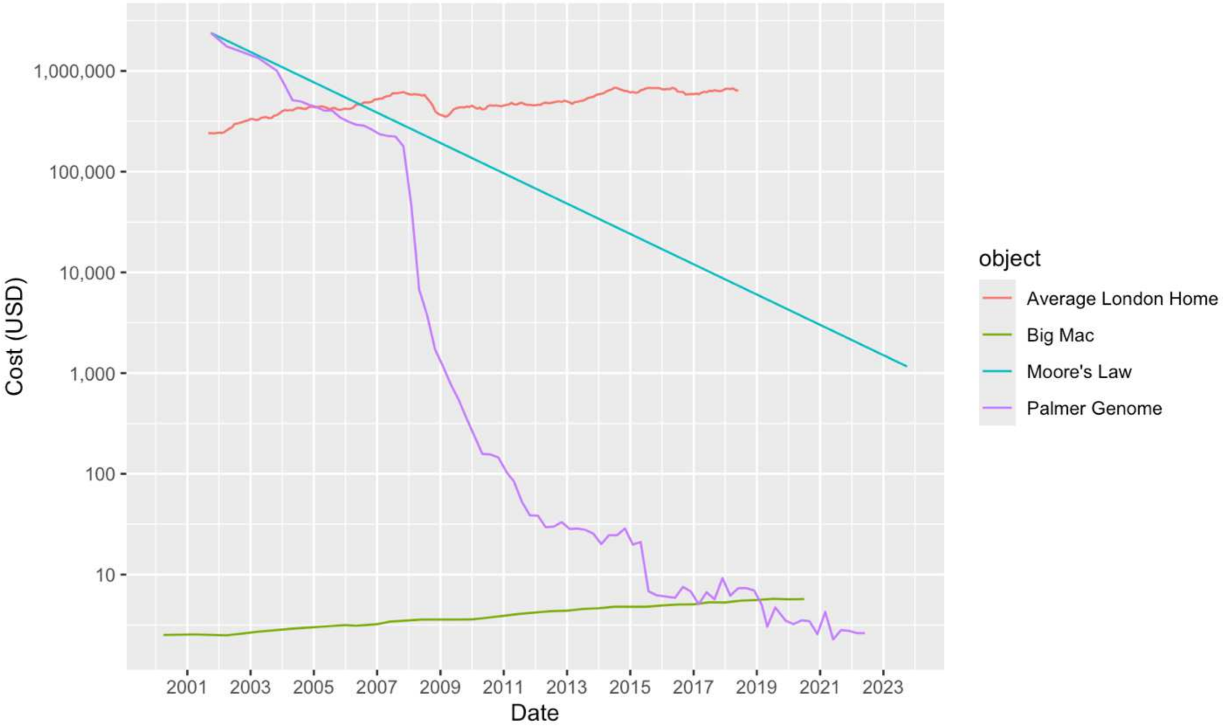


Figure 2-1. Price, over time, of a Big Mac® sandwich, the average home sold in London, and sequencing required to cover the genome of Palmer amaranth. Moore's law tracks the expected halving of sequencing cost every two years. Prices have been converted to USD and data collected from www.kaggle.com, www.landregistry.data.gov.uk, and www.genome.gov, respectively.

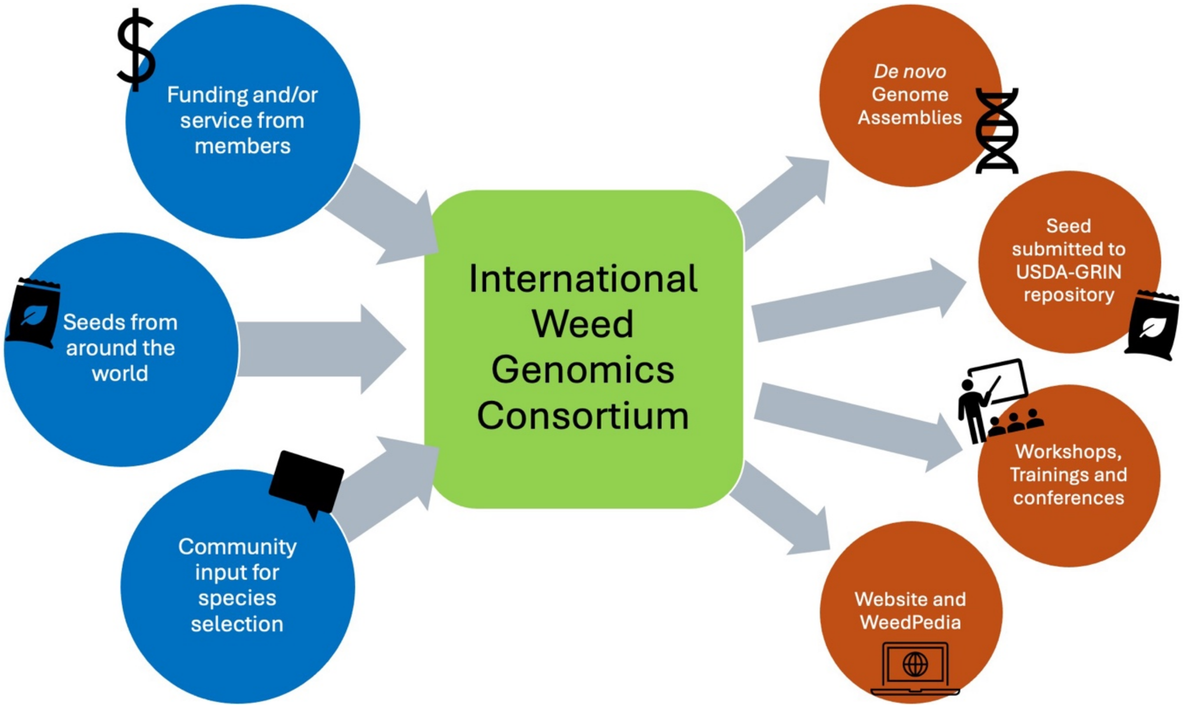


Figure 2-2. Schematic summarizing the inputs and outputs of the International Weed Genomics Consortium.

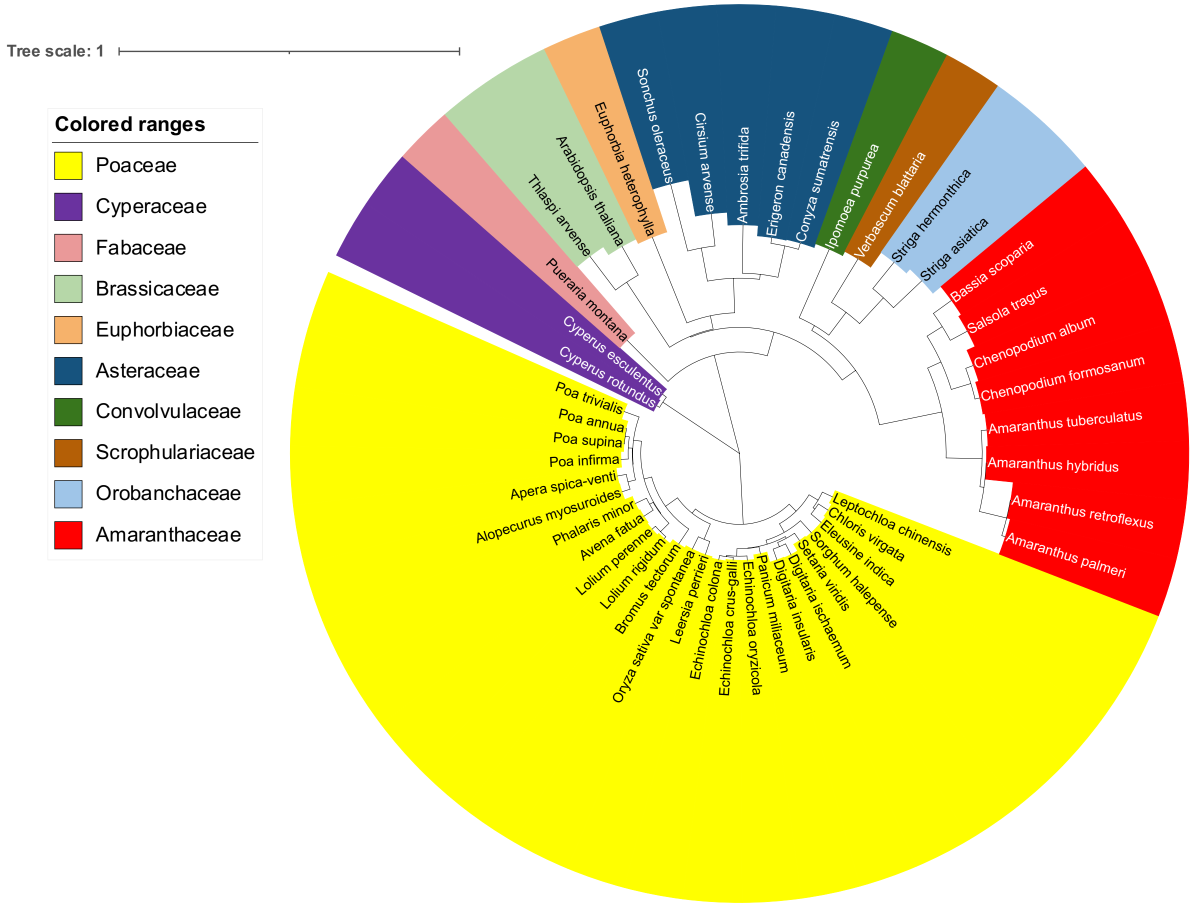


Figure 2-3. Phylogenetic tree showing relatedness of all weed species hosted on the WeedPedia database (as of May 23rd, 2024). Sequence of the *acetolactate synthase* gene was used to generate the neighbor joining tree and plant families are represented by different background colors. Sequence used to make the tree is available upon request from the authors.

References

Abramson J, Adler J, Dunger J et al. (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, https://doi.org/10.1038/s41586-024-07487-w

Bennetzen J, Schmutz J, Wang H *et al*. (2012) Reference genome sequence of the model plant *Setaria*. *Nat Biotechnol*, **30**, 555–561.

Bobadilla LK and Tranel PJ (2024) Predicting the unpredictable: the regulatory nature and promiscuity of herbicide cross resistance. *Pest Manag Sci*, **80**, 235-244.

Bolognesi R, Ramaseshadri R, Anderson J, *et al*. (2012) Characterizing the mechanism of action of double-stranded RNA activity against western corn rootworm (*Diabrotica virgifera virgifera LeConte*). *PLoS ONE*, **7**, e47534.

Brusa A, Patterson EL, Gaines TA, Dorn K, Westra P, Sparks CD, Wyse D. A needle in a seedstack: an improved method for detection of rare alleles in bulk seed testing through KASP. *Pest Manag Sci*, **77**, 2477-2484.

Gaines TA, Slavov GT, Hughes D, Küpper A, Sparks CD, Oliva J, Vila-Aiub MM, Garcia MA, Merotto A Jr, Neve P (2021) Investigating the origins and evolution of a glyphosate-resistant weed invasion in South America. *Mol Ecol*, **30**, 5360–5372.

Golebiowska-Paluch G, Dyda M (2023) The genome regions associated with abiotic and biotic stress tolerance, as well as other important breeding traits in triticale. *Plants (Basal)*, **12**, 619.

Kreiner JM, Latorre SM, Burbano HA, *et al*. (2022) Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science*, **378**, 1079-85.

Montgomery JS, Morran S, MacGregor DR *et al*. (2024) Current status of community resources and priorities for weed genomics research. *Genome Biol*, **25**, 139.

Laforest M, Martin SL, Bisaillon K, Soufiane B, Meloche S, Page E (2020) A chromosome-scale draft sequence of the Canada fleabane genome. *Pest Manag Sci*, **76**, 2158-2169.

Leon RG, Bassham DC (2024) PROTAC for agriculture: learning from human medicine to generate new biotechnological weed control solutions. *Pest Manag Sci*, **80**, 262-266.

Li W, Wang D, Jin T, Chang Q, Yin D, Xu S, Liu B, Liu L (2010) The vacuolar Na+/H+ antiporter gene SsNHX1 from the halophyte *Salsola soda* confers salt tolerance in transgenic alfalfa (*Medicago sativa* L.). *Plant Mol Biol Rep*, **29**, 278–290.

Lu H, Zhou Q, He J, Jiang Z, Peng C, Tong R, et al. (2020) Recent advances in the development of protein–protein interactions modulators: mechanisms and clinical trials. *Signal Transduct Target Ther*, **5**, 213.

Onkokesung N, Brazier-Hicks M, Tetard-Jones C, Bentham A, Edwards R (2022) Molecular diagnostics for real-time determination of herbicide resistance in wild grasses. *J Biotech*, **358**, 64-66.

Ravet K, Patterson EL, Krähmer H *et al*. (2018) The power and potential of genomics in weed biology and management. *Pest Manag Sci*, **74**, 2216-2225.

Riggins CW, Peng Y, Stewart Jr. CN, Tranel PJ (2010) Characterization of de novo transcriptome for waterhemp (*Amaranthus tuberculatus*) using GS-FLX 454 pyrosequencing and its application for studies of herbicide target-site genes. *Pest Manag Sci*, **66**, 1042–1052.

Rigon CAG, Küpper A, Sparks C *et al*. (2023) Cytochrome P450 CYP72A219 is involved in evolution of metabolic resistance to tembotrione in *Amaranthus palmeri* populations. *bioRxiv* 2023.12.13.571468. doi: https://doi.org/10.1101/2023.12.13.571468

Rodrigues TB, Mishra SK, Sridharan K et al. (2021) First sprayable double-stranded RNA-based biopesticide product targets proteasome subunit beta type-5 in Colorado potato beetle (*Leptinotarsa decemlineata*). *Front Plant Sci*, **12**. doi: 10.3389/fpls.2021.728652.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**, 796–815.

Vignieri S (2023) Zoonomia. *Science*, **380**, 356-357.

Zabala-Pardo D, Gaines T, Lamego FP, Avila LA (2022) RNAi as a tool for weed management: challenges and opportunities. *Adv Weed Sci*, **40**, e020220096.

CHAPTER II - A TRANSPOSABLE ELEMENT INSERTION IN *AUX/IAA16* INTERRUPTS NORMAL SPLICING AND GENERATES A NOVEL DICAMBA RESISTANCE ALLELE IN *BASSIA SCOPARIA*

Summary

A dicamba-resistant population of kochia (*Bassia scoparia*) previously identified in Colorado, USA in 2012 was used to generate a synthetic mapping population that segregated simply for dicamba resistance. Linkage mapping to associate injury following dicamba application with genotype information from restriction-site-associated DNA sequencing identified a single locus in the kochia genome associated with resistance on chromosome 4. A mutant version of *Auxin/Indole-3-Acetic Acid 16* (*AUX/IAA16*; a gene previously implicated in dicamba resistance in kochia) is found near the middle of this locus in resistant plants. Long read sequencing of resistant plants identified a recently inserted Ty1/Copia retrotransposon near the beginning of exon two of *AUX/IAA16*, leading to interruption of normal splicing. Stable transgenic lines of *Arabidopsis thaliana* ectopically expressing the mutant and wildtype alleles of *AUX/IAA16* were developed. *Arabidopsis thaliana* plants expressing the mutant *AUX/IAA16* allele grew shorter roots on control media but had less reduction in root growth on media containing either dicamba (5 µM) or IAA (0.5 µM) compared to non-transgenic plants or those expressing the wildtype allele of *AUX/IAA16.* Protein modeling suggests the substitution of a specific glycine residue in the degron domain of AUX/IAA16 is especially important for resistance. A fitness cost associated with the mutant allele of *AUX/IAA16* has implications for resistance evolution and management of kochia populations with this resistance mechanism. We report a molecular assay

for rapid detection of the mutant *AUX/IAA16* allele that can inform in-season management decisions of agricultural producers.

Introduction

Weed control in modern agricultural production systems largely rely on the application of herbicides. The first synthetic herbicides to be commercialized (like 2,4-D; described by Quastel, 1950) were eventually determined to mimic the endogenous plant hormone auxin. The low cost of these synthetic auxin herbicides matched with their ability to selectively control broadleaf weeds in cereal production settings led to rapid adoption of spray-based weed management strategies. As new herbicides were discovered and commercialized, mixing multiple herbicides in a single spray mixture allowed for robust control of many of the most important weeds in diverse agricultural settings. While evolution of herbicide resistant weed populations has reduced the efficacy of many herbicides, evolution of resistance to synthetic auxin herbicides has been relatively slower and less frequent than other modes of action (Busi et al., 2018; Heap, 2024). Thus, some of the oldest herbicides (e.g., 2,4-D and dicamba) remain some of the most widely used across many agricultural settings. In fact, synthetic auxin resistance traits for dicot crops were recently developed and commercialized to provide producers more options to control weed populations that had evolved glyphosate resistance (Behrens et al., 2007; Wright et al., 2010). These two traits expanded the potential use cases and value of two of the most used synthetic auxin herbicides, dicamba and 2,4-D. This expansion in use for these two herbicides has had the unintended effect of increasing selection pressure for resistance alleles and has correlated with a recent increase in the number of resistance reports to these herbicides. Beyond weed control, these synthetic auxin compounds have been crucial for experiments to understand auxin perception and signaling, perhaps the most complex of all plant signaling pathways (Ma et al., 2017).

While the physiological mechanism of action for many herbicides have been well understood for decades, the cascade of cellular responses following the application of a synthetic auxin herbicide is convoluted and is the subject of many recent studies. The topic has been reviewed several times, with new experimental results incrementally advancing our knowledge of the key factors leading to plant death (Grossmann, 2010; Song, 2014; Christoffoleti et al., 2015; McCauley et al., 2020). Auxin, whether synthetic or natural, first binds to one of the isoforms of Transport Inhibitor Response 1/Auxin-Signaling F-Box (TIR1/AFB) (Dharmasiri et al., 2005; Kepinski & Leyser, 2005). Once bound to TIR1/AFB, auxin acts as a “molecular glue”, mediating the interaction between TIR1/AFB and Auxin/Indole-3-Acetic Acid (Aux/IAA) proteins. Identification of conserved domains and experimentation have produced a list of residues in TIR1/AFB and Aux/IAA that are vital for this interaction (Tan et al., 2007; Uzunova et al., 2016; Ramans-Harborough et al., 2023). This interaction leads to the ubiquitination and eventual degradation of Aux/IAA proteins (Gray et al., 2001; Zenser et al., 2001). Because Aux/IAA proteins act as transcriptional repressors of auxin response factor (ARF) proteins, Aux/IAA degradation leads to the modulation of auxin-responsive gene expression (Grossmann, 2010). TIR1/AFB and Aux/IAA each have several orthologs, allowing for diversification for different isoforms. For example, LeClere et al. (2018) show that Aux/IAA16 preferentially binds with TIR1 and AFB6 (a homolog of AFB found in members of amaranthaceae, but not brassicaceae) in the presence of some auxin analogs, but not others. Different synthetic auxin herbicides likely promote the interaction of certain Aux/IAA proteins with certain TIR1/AFB proteins more efficiently, but this network of interactions is still not well understood.

A disruption in any step of this auxin signaling system (e.g. TIR1 binding auxin, Aux/IAA interacting with TIR1/AFB, or the ubiquitination and degradation pathway of Aux/IAA) could cause reduced activity of synthetic auxin herbicides (Ruegger et al., 1998; LeClere et al., 2018; Todd et al., 2020). Thus, evolution of synthetic auxin resistance should be relatively simple and common across weed species in the face of incredibly pervasive and strong selection pressure. However, mutants causing resistance to synthetic auxin herbicides often also cause a reduction in fitness compared to the wildtype (Roux & Reboud, 2005; Wu et al., 2021). While many auxin resistance alleles have likely evolved in many species, these alleles were likely rapidly purged during temporal pauses in management-induced selection for those alleles (i.e., if synthetic auxin herbicides were not used for several generations). Nevertheless, there are several reported cases of synthetic auxin herbicide resistance evolving and becoming widespread in weed species. Interestingly, all but one of these reports that specify a specific genetic basis (de Figueiredo et al., 2022a) pertain to mutations in or near the degron domain of Aux/IAA genes. LeClere et al. (2018) and Ghanizadeh et al. (2024) report, in closely related species, different mutations of a specific glycine residue in AUX/IAA16 that is known to be vital to normal auxin signaling. In an unrelated species, de Figueiredo et al. (2022b) report a 27-bp deletion that shortens the distance between the degron and PB1 domains of IAA2.

Here, we seek to understand the genetic and molecular basis of dicamba resistance in a population of kochia (*Bassia scoparia*) collected from Colorado, USA. We use linkage mapping in a population segregating for resistance to identify a region of the genome that segregates with resistance and contains a mutant version of *AUX/IAA16*. Long read sequencing allowed us to identify a transposable element insertion in the coding region of *AUX/IAA16* that disrupts normal splicing and changes several amino acids in the protein product relative to the wildtype. Ectopic expression in *Arabidopsis thaliana* proves this mutant *AUX/IAA16* is sufficient to cause resistance to dicamba, and protein modeling points to the importance of a glycine-to-threonine substitution in the degron domain. We finish with a discussion of a fitness cost associated with dicamba resistance and implications towards agricultural management of dicamba-resistant weed populations.

Results and Discussion

Characterization of herbicide resistance

Dose response experiments were conducted to confirm and quantify the level of dicamba and 2,4-D resistance in a population of kochia (named M32) collected from Akron, Colorado, USA (Figures 3-1A and 3-1B). Parameter estimates for two-parameter log-logistic models for each population used in this experiment are shown in Table 3-1 and Table 3-2. In the dicamba experiment, populations M32 and 9425 (a known dicamba-resistant kochia population previously characterized by LeClere et al. (2018) and Pettinga et al. (2018)) had ED50 estimates 7.6 ($p<0.0001$) and 10.5 ($p<0.0001$) times greater than an herbicide-sensitive kochia population (7710). The ED50 parameter represents the dose of dicamba needed to cause 50% injury to that specific population. These estimates were significantly different from the herbicide-sensitive control and from each other ($p<0.0001$), indicating dicamba resistance in M32 is not as high as in 9425. Nevertheless, the ED50 estimate for both resistant populations were greater than the recommended field rate of dicamba (560g ha^{-1}), suggesting both populations would not be controlled by dicamba in a field setting. Similarly, in the 2,4-D experiment, the ED50 estimate for M32 was 5.0 ($p<0.0001$) times greater than that of the herbicide-sensitive control and greater than the recommended field rate for 2,4-D.

While herbicide-sensitive plants show epinasty starting approximately 5 days after treatment with dicamba or 2,4-D, resistant plants in our study do not show symptoms at all when treated with a field rate of either herbicide. This lack of initial injury suggests that resistant plants either rapidly metabolize/sequester the herbicides before they cause injury (Behrens et al., 2007; Ge et al., 2010) or possess mutant target site(s) that reduce their interaction with their molecular target(s) (Walsh et al., 2006; LeClere et al., 2018; de Figueiredo et al., 2022b). However, in documented cases of metabolic auxin herbicide resistance, resistant plants do show symptoms of herbicide damage but eventually recover (de Figueiredo et al., 2022a; Todd et al., 2024). It is true that highly efficient metabolism genes from bacteria protects plants from herbicide-induced symptoms (Behrens et al., 2007; Wright et al., 2010), but that has not been the case for weeds with evolved auxin herbicide resistance.

To understand if physiological mechanisms keep dicamba from being absorbed or translocated within resistant plants, radiolabeled dicamba was used to track its movement in dicamba-resistant and -sensitive plants. Significantly less dicamba was absorbed by herbicide-sensitive plants over 192 h compared to the two resistant populations, but rate of absorption (time required for 90% absorption) was not significantly different between populations (Figures 3-1C, Table 3-3). Therefore, reduced dicamba absorption is not responsible for resistance. Time series data for the translocation of radiolabeled dicamba out of the treated leaf did not fit well to any tested models, so t-tests were used to determine if means differed between resistant and sensitive populations at each time point (Figure 3-1D). Dicamba-resistant plants translocated less dicamba out of the treated leaf compared to the sensitive plants. These results suggest that the dicamba resistance mechanism in the experimental population is physiologically similar to one described by Pettinga et al. (2018) and LeClere et al. (2018) that is caused by a mutation in *BsAUX/IAA16*.

A single locus on chromosome 4 is associated with dicamba resistance

A quantitative trait loci (QTL) mapping approach was taken to identify regions of the genome associated with dicamba resistance. Segregating F_3 populations were developed from multiple independent biparental crosses using a dicamba-resistant plant as the male parent and a dicamba-sensitive plant as the female parent. We expected to find one dominant locus controlling dicamba-resistance because the ratio of alive:dead F_3 plants following dicamba treatment at a rate of 560 g ha^{-1} (200:85) did not significantly differ from 3:1 ($p = 0.06$). We hypothesize that the over-abundance of dead (herbicide-sensitive) plants may be caused by a fitness penalty associated with the dicamba resistance mechanism that reduces germination frequency (see section on fitness cost below). A novel double-digest restriction site associate DNA sequencing (ddRADseq) protocol was used to determine genotypes of 103 segregating F_3 plants for 1,592 variant loci across the genome. A genome scan associating this genotype information to visual injury rating identified one significant QTL that peaks at 88 Mbp on chromosome 4 (Figure 3-2A). This interval contains *BsAUX/IAA16* at 87.4 Mbp, a gene that is important for the signaling of auxin and has been implicated in auxin herbicide resistance in kochia and a closely related species (LeClere et al., 2018; Ghanizadeh et al., 2024). The genotype of *BsAUX/IAA16* within this QTL of plants from an independent segregating F_3 line explained 43% of variation in visual injury following dicamba treatment at 560 g ha^{-1} (Figure 3-2B). While survival at this rate of dicamba may be dominant, visual injury is incompletely dominant (Figure 3-2B) with a calculated degree of dominance of -0.58 (Falconer, 1964). This dominance has important implications for the spread of herbicide resistance alleles. For instance, a dominant trait is likely to spread more quickly throughout a population, but is less likely to be driven to fixation under selection (Holmes et al., 2022).

A transposable element insertion in BsAUX/IAA16 causes a change in exon splicing

Sequencing transcripts of *BsAUX/IAA16* from plants homozygous for either either parental allele of the QTL discovered on chromosome 4 revealed near identical translated protein sequences except for the region around the degron domain (Figure 3-3B; Table 3-4). The degron of *BsAUX/IAA16* is located near the splice junction of exons one and two (Figure 3-3A). To understand the sequence of genomic DNA in and around *BsAUX/IAA16*, a plant homozygous for the resistant parent allele of *BsAUX/IAA16* was used for PacBio HiFi sequencing. After *de novo* assembly, we discovered a ~3.5 Kbp insertion at the beginning of exon two of *BsAUX/IAA16* (Figure 3-3A). The insertion disrupts splicing of *BsAUX/IAA16*, replacing 16 bases at the beginning of exon two in the wildtype with the final 19 bases of the insertion in the mutant (Figure 3-3A). The mutant allele will herein be reffered to as $\textit{BsAUX/IAA16}_{Mut}$, while the wildtype allele will be referred to as $\textit{BsAUX/IAA16}_{WT}$. This results in the addition of one amino acid and the replacement of four more in the final protein product (Figure 3-3B; Table 3-4). The new splice sites contain canonical GT and AG motifs. This change includes a substitution of a glycine to threonine at the N-terminal end of degron domain (Figure 3-3B; Table 3-4). LeClere et al. (2018) showed that a substitution of this glycine to asparagine in BsAUX/IAA16 is sufficient for dicamba resistance in kochia. Similarly, Ghanizadeh et al. (2024) associate a substitution of this glycine to aspartic acid with dicamba resistance in *Chenopodium album*, a closely related species. Finally, NMR experiments by Ramans-Harborough et al. (2023) show that this glycine is involved in the interaction between IAA and TIR1. Given the demonstrated importance of this glycine residue, we hypothesize that the glycine to threonine substitution is the major driver of dicamba resistance. Protein modeling shows that the larger side group of threonine contacts the surface of the TIR1 binding pocket and causes steric hindrance of the IAA/TIR1 interaction (Figure 3-3C). The rest of the mutant residues are found outside of the binding pocket of TIR1 and in a disordered domain; thus, they are unlikely to affect IAA/TIR1 interactions (Tan et al., 2007; Ramans-Harborough et al., 2023).

The retrotransposon insertion in *BsAUX/IAA16*$_{Mut}$ has 5 bp target site duplications flanking 429 bp long terminal repeats (LTRs), identifying it as a Class I LTR element (Wicker et al., 2007). No protein coding sequences were found within this insertion, suggesting this element is non-autonomous and relies on another transposable element for transposition. Aligning the insertion to the *de novo* assembly built from a dicamba-resistant plant and the reference genome built from a herbicide-sensitive plant identifies a near perfect match (99.3% identity) 1 Mbp away on chromosome 4 and another match (97% identity) on chromosome 2 in both assemblies (Table 3-5). Because the site on chromosome 4 is the most similar across the genome and retrotransposons use a "copy and paste" mechanism of transposition, we hypothesize this to be the donor location for the insertion in *BsAUX/IAA16*$_{Mut}$. A maximum likelihood phylogenetic tree supports this idea, as the chromosome 4 sequences from both assemblies form a distinct clade from those on chromosome 2 (Figure 3-4). When comparing the insertion in *BsAUX/IAA16*$_{Mut}$ with its donor site on chromosome 4, there is one single nucleotide polymorphism (SNP) and two single base pair insertions (Table 3-5). These insertions are found in homopolymer runs of at least 9 bp and are likely artifacts of the sequencing technology (Espinosa et al., 2024). Given the approximated rate of mutation in *Arabidopsis thaliana* of 7×10^{-9} base substitutions per site per generation (Ossowski et al., 2010), we estimate a new single base substitution to be generated in either one of these 3,466 bp regions approximately every ~20,000 generations. This estimation is a very crude and extremely young in the context of evolutionary time. Future research using this insertion could be used to test the canonical paradigm that weedy adaptation is the consequence of selection on standing genetic variation versus the more contentious proposition that management-induced stress can generate contemporary adaptive alleles in weeds (Dyer, 2018).

To identify autonomous version(s) of the retrotransposon within *BsAUX/IAA16*$_{Mut}$, the LTR sequence was aligned to the 7710 reference genome (Hall et al., 2023). Putative matches were filtered by sequence similarity, distance between LTRs, orientation of the LTRs, and containment of coding sequences necessary for transposition. This process led to the identification of a single autonomous element, named gi_5_13kb (Figure 3-5). A phylogenetic tree built from the LTRs of known classes of retrotransposons suggest this to be a Ty1/Copia element (Figure 3-6) that we have named *Outlaw*. This one version of Outlaw seems to be driving transposition of this retroelement in the kochia genome.

BsAUX/IAA16$_{Mut}$ *expression is sufficient for dicamba resistance in Arabidopsis thaliana*

To demonstrate the role of *BsAUX/IAA16*$_{Mut}$ in dicamba resistance, stable transgenic lines of *Arabidopsis thaliana* expressing *BsAUX/IAA16*$_{WT}$ or *BsAUX/IAA16*$_{Mut}$ were generated through the floral dip method. When seeds were grown on phytoagar plates without herbicide, Arabidopsis expressing *BsAUX/IAA16*$_{Mut}$ grew shorter roots compared to those expressing *BsAUX/IAA16*$_{WT}$ or with no transgene (Figure 3-7A). This is further evidence of a likely fitness cost associated with disrupting normal auxin signaling in roots as seen by de Figueiredo et al. (2022b) and LeClere et al. (2018). When grown on media containing 5 µM dicamba, only seedlings expressing *BsAUX/IAA16*$_{Mut}$ were able to grow and had significantly longer roots when normalized to the length of roots grown on solvent-only media (Figure 3-7B). A similar result was seen when seedlings were grown on media containing 0.5 µM IAA (Figure 3-7C), suggesting this variant of AUX/IAA16 also reduces the perception of natural auxin. No seeds grew roots on media containing 5 µM or 0.5 µM 2,4-D.

To confirm the results of root growth assays that indicate *BsAUX/IAA16*$_{Mut}$ is sufficient for dicamba resistance, plants of select T$_3$ lines either expressing *BsAUX/IAA16*$_{WT}$, *BsAUX/IAA16*$_{Mut}$, or no transgene were grown in the greenhouse and treated with 140 g dicamba ha^{-1}. Only one of the lines expressing *BsAUX/IAA16*$_{Mut}$ showed less visual injury than the other lines tested and appeared stunted and darker green than the other lines even when no dicamba treatment was administered (Figure 3-8). Quantitative PCR indicated similar expression levels of *BsAUX/IAA16* between lines expressing *BsAUX/IAA16*$_{WT}$ or *BsAUX/IAA16*$_{Mut}$ (Figure 3-7D). However, plants expressing *BsAUX/IAA16*$_{Mut}$ had no detected induction of auxin response genes *AtIAA19* or *AtGH3.3* six hours after dicamba treatment (Figure 3-7E and 3-7F). These genes are known to have increased expression following exposure to natural or synthetic auxins (Tatematsu et al., 2004; Gleason et al., 2011; de Figueiredo et al., 2022b). Differences between root and whole plant assays may be caused by differences in expression of the transgene between root and above ground tissue.

The BsAUX/IAA16$_{Mut}$ *allele is associated with a reduction in fitness*

Because we hypothesize this insertion interferes with natural auxin signaling as shown by LeClere et al (2018) and de Figuieredo et al. (2022), we expect this change to also confer a fitness penalty in the absence of synthetic auxin application. To formally test this, kochia plants were grown from an F_3 family that is segregating for the *BsAUX/IAA16* locus and fitness traits were measured without the application of auxin. The ratio of genotypes significantly differed from the expected 1:2:1 ($p<0.0001$), with fewer than expected plants with the mutant *BsAUX/IAA16* locus (Figure 3-8). All seedlings that emerged were used in the experiment, so we hypothesize that $\textit{BsAUX/IAA16}_{Mut}$ causes reduced germination. Further, $\textit{BsAUX/IAA16}_{Mut}$ was associated with shorter plants, both at transplanting and at maturity (Figure 3-8). Evidence of a fitness cost associated with herbicide resistance has implications on kochia management. For instance, if dicamba is not used for several years, the frequency of $\textit{BsAUX/IAA16}_{Mut}$ is expected to decrease. High gene flow in kochia facilitates rapid spreading of alleles across long distances and maintenance of high genetic diversity across the genome following selection of rare herbicide resistance alleles (Martin et al., 2020; Ravet et al., 2021). The negative pleotropic effect of $\textit{BsAUX/IAA16}_{Mut}$ may help limit the expected spread of this resistance allele and allow producers to better manage resistant populations that have already established.

Taken together, our results further emphasize the importance of the glycine residue within the degron domain of AUX/IAA proteins in auxin signaling. They also reinforce the importance of *BsAUX/IAA16*, specifically, in dicamba action. We describe the first instance of the insertion of a mobile genetic element creating an herbicide resistance allele through gain-of-function. These results highlight the amazing adaptability of weedy genomes, and the unlikely ways genomes evolve in the face of strong selection pressures (Montgomery et al., 2024).

Materials and Methods

Plant growth and herbicide treatment conditions

Unless otherwise noted, all plants were grown in greenhouses at Colorado State University in 3.8-cm by 3.8-cm by 5.8-cm pots containing fine-grade potting mix (Fafard #2-SV; American Clay Works, Denver, CO). Temperatures were held between 22C and 24C, and supplemental light was provided by liquid halogen grow lights to ensure a photoperiod of 14h/10h. Herbicide applications were made using a moving overhead single-nozzle sprayer (DeVries Manufacturing, Hollandale, MN) calibrated to deliver 187 L ha^{-1}. Engenia (BASF, Research Triangle Park, NC) was used for dicamba applications, and Clean Amine (Loveland Products, Inc., Loveland, CO) was used for 2,4-D applications. For kochia, herbicide applications took place when plants were 10-15 cm tall, and for Arabidopsis, plants were sprayed when they had 10-12 leaves and before bolting.

Plant material

The M32 population of kochia was collected near Akron, Colorado as part of an herbicide resistance survey conducted between 2012 and 2014 (Westra et al., 2019). Seeds from this field collection were grown and the resulting plants were treated with 560 g dicamba ha^{-1}. Survivors were open-pollinated, and their seed was combined to form a composite dicamba-resistant seed lot. Approximately 50 seeds from this combined lot were grown and treated with 560 dicamba ha^{-1}. The individual with the least herbicide injury was used as a male parent in a biparental cross with an individual from the population 7710, an inbred, herbicide-susceptible population (Preston et al., 2009) that was used to generate the first reference genome assembly for kochia (Patterson et al., 2019). To make the cross, immature flowers of the dicamba-sensitive plant were emasculated using forceps under a microscope, and the emasculated flowers were labeled. The two plants were then grown to maturity together in a pollen-exclusion tent. Seeds from the dicamba-sensitive plant were collected and grown; the resulting plants were treated with 560 g dicamba ha^{-1}. A Kompetative Allele Specific PCR (KASP) assay that detects *BsAUX/IAA16*$_{Mut}$ was used to confirm survivors as hybrids (primer sequence and thermocycler conditions found in Table 3-6). Several confirmed hybrid plants were grown in pollen exclusion tents and self-pollinated to create several F_2 families. Plants from two F_2 families were grown, but because of low seed production in the F_1 generation, plants were allowed to self-pollinate with no selection to produce F_3 families.

Dose response

Plants were grown from the 7710 (Preston et al., 2009) and 9425 (LeClere et al., 2018) populations as well as from the M32 composite resistant seed lot. Plants were treated with one of several rates of dicamba. These rates included 0, 8.75, 17.5, 35, 70, 140, 280, 560, and 1120 g dicamba ha^{-1} for the herbicide-sensitive population and 0, 70, 140, 280, 560, 1120, 2240, and 4480 g dicamba ha^{-1} for the herbicide-resistant populations. Six uniform plants were used for each population at each herbicide rate before being returned to the greenhouse. Visual injury and survival were rated 21 days after treatment (DAT). The drc package (v3.0-1; (Ritz et al., 2015)) in R (v4.0.2; (R-Core-Team, 2021)) was used to fit the injury data to a 2-parameter log-logistic dose response curve. Significant difference is in the ED50 parameter were determined using the compParm function from the drc package (Ritz et al., 2015).

Herbicide absorption and translocation profiles

Kochia plants from M32, 7710, and 9425 were germinated in a growth chamber (60% relative humidity, 21/18 C, and 16/8 h photoperiod) in potting soil and transplanted to fine sand when they reached ~3 true leaves. The plants were irrigated with fertilizer until the plants reached 10 cm in height. At this point, aluminum foil was used to cover the second youngest fully expanded leaf while the plants were sprayed with 560 g dicamba ha^{-1}. After spraying, the aluminum foil was removed, the leaf was then marked and treated with 10 µL of a ^{14}C-labeled dicamba solution (total radioactivity of 3.33 KBq or 200,000 dpm per plant). Plants were returned to the growth chamber until sampling at 3, 6, 12, 24, 48, 96, and 192 h after treatment. The treated leaf, remaining above ground tissue, and below ground tissue of three biological replicates were separated for each population at each time point. The treated leaf was washed in 10 mL 10% methanol + 1% NIS, and radioactivity in this wash solution was quantified in 10 mL of scintillation mixture (Ecoscint XR, National Diagnostics, Atlanta, GA) using liquid scintillation spectrometry (Packard Tricarb 2300TR, Packard Instrument Co., Meriden, CT). Plant tissue was dried in an oven at 60 C for at least 14 d before oxidation in a biological oxidizer (OX500; RJ Harvey Instrument Co., Tappan, NY) followed by radioactivity measurement by liquid scintillation spectrometry. Absorption time series data was fitted to a rectangular hyperbolic model, and the parameters “absorption max” (*Amax*) and “time to 90 percent absorption” (*t90*) were estimated and compared in R (Kniss et al., 2011; Ritz et al., 2015; R-Core-Team, 2021). Because the time series data for translocation of dicamba outside of the treated leaf did not fit any of the models proposed by Kniss et al. (2011), t-tests were used to determine if the amount of dicamba retained in the treated leaf was different between dicamba-resistant and dicamba-sensitive populations at each time point.

QTL mapping

Plants from several F_3 families derived from a biparental cross described above were grown and treated with 560 g dicamba ha^{-1}. Two families that segregated for survival following dicamba treatment (4-1-1 and 4-5-10, each from a different F_1 hybridization event) were used for QTL mapping. Approximately 150 plants from the F_3 family 4-5-10 were grown, and a single young leaf was sampled and frozen in liquid nitrogen before treatment with 560 g dicamba ha^{-1}. Visual injury was rated for each plant 21 DAT, and DNA was extracted from each plant (including the original parents of the cross) using the CTAB method (Doyle & Doyle, 1990). The concentration of each DNA sample was quantified using the Qubit™ dsDNA BR kit (Thermo Scientific, Waltham, MA) and diluted to 20 ng/µl. A double-digest restriction-site associated DNA sequencing (ddRADseq) protocol was developed in conjunction with the University of Minnesota Genomics Center (Minneapolis, MN) by estimating the number of RAD sites across the reference genome (Patterson et al., 2019) for several enzyme combinations and then empirically testing the BtgI-TaqI combination on several samples at multiple sequencing depths. DNA from the two original parents and 103 plants from the 4-5-10 F_3 family were used for ddRADseq with BtgI and TaqI as the restriction enzymes with a target of 4 million 150 bp single end reads per sample. Library prep and sequencing was completed by the University of Minnesota Genomics Center.

Reads from each sample were passed through a variant calling pipeline that can be found at github.com/JMont12/dicamba_kochia. Briefly, raw reads were trimmed with Trimmomatic (v0.36; (Bolger et al., 2014)) and aligned to the reference genome of kochia (Hall et al., 2023) using the burrows-wheeler aligner (v0.7.17, (Li & Durbin, 2009)). The alignments were processed with samtools (v1.15, (Li et al., 2009)) and passed to GATK (v4.2.0, (Poplin et al., 2018; Van der Auwera & O'Connor, 2020)) to call variants and filter them based on depth, quality, and strand bias. Variants were further filtered using custom scripts to only include biallelic single nucleotide polymorphisms that were homozygous and different between the two original parents. Filtered variants and phenotype data were used to conduct a genome scan within the qtl2 package in R (v0.32, (Broman et al., 2019)). Because a permutation test for significance was not possible, a significance threshold was established through a Bonferroni correction of alpha=0.05 considering the number of tests to equal the number of filtered markers used in the scan then converting the adjusted p-value to LOD as described by Nyholt (2000). The genotype around the QTL discovered on chr4 was determined for 210 plants from another F_3 family (4-1-1) using the KASP assay developed for the *BsAUX/IAA16* gene (primer sequence and thermocycler conditions in Table 3-6). A simple linear model was fitted using visual injury 21 DAT as the dependent variable and *BsAUX/IAA16* genotype as the independent variable and plotted using ggplot2 (v3.4.4, (Wickham, 2016)) in R (R-Core-Team, 2021).

BsAUX/IAA16$_{Mut}$ *sequencing*

Total RNA was extracted using the Direct-zol RNA Microprep kit (Zymo Research, Irvine, CA) from individuals of the 4-5-10 selected line that were homozygous for either parental allele of *BsAUX/IAA16* based on the results of KASP testing. This RNA was used to generate cDNA libraries for each sample using the ProtoScript® II First Strand cDNA Synthesis Kit (New England Biolabs, Ipswich, MA). Primers that bind to the beginning and end of the *BsAUX/IAA16* gene (Table 3-7) were used with the EconoTaq PLUS Green kit (LGC, Teddington, Middlesex, UK) to amplify the full-length coding sequence. The product of each PCR reaction was run on a 1% agarose gel to ensure a single product, then sent to Azenta (Burlington, MA) for Sanger sequencing using the forward and reverse primers.

To determine the genomic sequence around the *BsAUX/IAA16*$_{Mut}$, 2 g of dark-treated leaf tissue from a plant that was homozygous for *BsAUX/IAA16*$_{Mut}$ was sampled and flash frozen in liquid nitrogen. This tissue was sent to Corteva Agriscience Center for Genome Excellence (Johnston, IA) for whole genome PacBio HiFi sequencing to a depth of ~30X. The resulting reads were assembled with HiFiasm (Cheng et al., 2021), and *BsAUX/IAA16* was located by BLAST (Camacho et al., 2009).

Characterization of transposable element insertion

To find similar elements to the one inserted in *BsAUX/IAA16*$_{Mut}$, the sequence of the 429 bp long terminal repeats (LTRs) of the insertion were aligned to the kochia 2.0 reference genome assembly (Hall et al., 2023) and to the newly created assembly described above using BLAST (Camacho et al., 2009). Filters to identify the autonomous version of the element included: two alignments of >85% identify and >100bp length must be <25 kb, but >4 kb apart and in the direct orientation and containing all required retrotransposon domains. Necessary retrotransposon coding sequences were identified manually and with TESorter (Zhang et al., 2022). To classify the retroelement, the autonomous version of the element was included in a phylogenetic tree with 25 complete *Arabidopsis thaliana* LTRs, representing each domain.

Development of transgenic Arabidopsis thaliana lines

The coding sequence of *BsAUX/IAA16*$_{Mut}$ and *BsAUX/IAA16*$_{WT}$ were amplified from cDNA libraries discussed above using the PrimeSTAR MAX DNA Polymerase kit (Takara Bio USA, Inc., San Jose, CA) and primers with tails that complement restriction sites in the expression vector *pFGC5941* (primer sequences and thermocycler conditions in Table 3-8). Empty *pFGC5941* was digested with AscI and BamHI (New England Biolabs, Ipswich, MA), and the *BsAUX/IAA16* amplicons were ligated into the digested vector using the In-Fusion Cloning kit (Takara Bio USA, Inc., San Jose, CA). Plasmid sequencing ensured correct assembly of expression vectors (Plasmidsaurus, Eugene, OR). Vectors containing *BsAUX/IAA16*$_{Mut}$ or *BsAUX/IAA16*$_{WT}$ were transformed into the GV3101 strain of *Agrobacterium tumefaciens*, and subsequently transfected into the Col-0 genotype of *Arabidopsis thaliana* using the floral dip method (Clough & Bent, 1998). Transformants were selected by spraying T_1 seedlings with BASTA herbicide (BASF, Research Triangle Park, NC) and confirmed with PCR for the transgene (primers and thermocycler conditions found in Table 3-9). Confirmed transformants were inbred for two generations, and T_3 populations fixed for the transgene were identified by spraying seedlings with BASTA herbicide and PCR for the transgene.

Effect of dicamba on transgenic Arabidopsis lines

Seeds of Col-0 and of fixed T_3 lines of Arabidopsis, each from a different transformation event, were gas sterilized and plated on media plates containing synthetic or natural auxin as described by de Figueiredo et al. (2022b). Plates were incubated in a growth chamber with a yellow light filter to prevent indole-3 acetic acid degradation. Ten seeds were used per population for each treatment. Root growth was measured 7 d after moving plates to growing conditions. T-tests were used to detect significant differences between lines expressing *BsAUX/IAA16*$_{Mut}$ compared to the Col-0 control. Results were plotted with ggplot2 in R.

Plants from select T_3 lines were grown to the 10-leaf stage and either untreated or treated with 140 g dicamba ha^{-1}. Four plants of each population were used for each treatment. Photos were taken 21 DAT to illustrate visual injury. Young leaf tissue was collected from treated and untreated plants 6 h after treatment. Total RNA was extracted from these leaf samples using the Direct-zol RNA Microprep kit (Zymo Research, Irvine, CA), and cDNA libraries were generated using the ProtoScript® II First Strand cDNA Synthesis Kit (New England Biolabs, Ipswich, MA). Following the methods of de Figueiredo et al. (2022b), relative expression was quantified for the *BsAUX/IAA16* transgene (primers and thermocycler conditions available in Table 3-10) and the auxin response genes *AtIAA19* and *AtGH3.3* using *AtCyclophilin* as a reference gene.

In silico mutagenesis of IAA variants

The crystal structure of the degron of IAA7 bound to TIR1 in the presence of 2,4-D (2p1n) was downloaded from the RCSB Protein Data Bank (Tan et al., 2007). Mutagenesis of the G127 residue was conducted and visualized in PyMOL.

Tables

Table 3-1. Parameter estimates for 2-parameter log-logistic dose response curves for M32, 9425, and 7710 kochia populations in response to dicamba. The equation fitted for each population was $y = \frac{100}{1+b(\log(x)-\log(ED50))}$ where y is predicted visual injury 21 days after dicamba treatment, x is the rate of dicamba in g ha^{-1}, *ED50* is the rate of dicamba required to cause 50% visual injury, and b is the slope of the curve at x=ED50.

Parameter	Estimate	Std. Error
Slope:7710	-2.9	0.3
Slope:M32	-4.5	0.9
Slope:9425	-2.1	0.3
ED50:7710	130.6	16.9
ED50:M32	988.3	27.2
ED50:9425	1370.4	20.9

Table 3-2. Parameter estimates for 2-parameter log-logistic dose response curves for M32 and 7710 kochia populations in response to 2,4-D. The equation fitted for each population was $y = \frac{100}{1+b(\log(x)-\log(ED50))}$ where y is predicted visual injury 21 days after dicamba treatment, x is the rate of dicamba in g ha^{-1}, *ED50* is the rate of dicamba required to cause 50% inury, and *b* is the slope of the curve at x=ED50.

Parameter	Estimate	Std. Error
Slope:7710	-1.5	0.1
Slope:M32	-2.3	0.4
ED50:7710	543.1	21.2
ED50:M32	2720.7	307.8

Table 3-3. Parameter estimates for rectangular hyperbolic models fitted to herbicide absorption data for M32, 9425, and 7710 kochia populations. The equation fitted for each population was $y = (Amax \times t)/(0.11 \times t90 + t)$ where y is percent of applied herbicide that was absorbed, t is time after herbicide application (in hours), $t90$ is the time required for 90% herbicide absorption, and $Amax$ is the maximum amount of herbicide absorbed.

Parameter	Estimate	Std. Error
Amax:7710	66.7930	5.6192
Amax:M32	64.4106	3.8674
Amax:9425	45.6638	5.3495
t90:7710	27.1771	9.4808
t90:M32	8.2880	4.0933
t90:9425	19.3150	14.6292

Table 3-4. CLUSTAL multiple protein sequence alignment of 7710 and M32 alleles of IAA16. The degron domain is bolded.

```
7710          MLSNERDKYTIDFEETELRLGLGLGIGLAGAADGDQLAKNNNGKRGFSETEGDSSVDLKL
M32           MLSNERDKYTIDFEETELRLGLGLGIGLAGAADGDQLAKNNNGKRGFSETEGDSSVDLKL
              ************************************************************

7710          NLSSSTTTTASTTTTNTTATKTTAENVKESKLDKSVNSGVDQKLKEKVASTTADPAKPTP
M32           NLSSSTTTTASTTTTNTTATKTTAENVKESKLDKSVNSGVDQKLKEKVASTTADPAKPTP
              ************************************************************

7710          AKT-QVVGWPPVRAFRKNIVAAHKKTSDDQTDQKASSNAITSAAFVKVSMDGAPYLRKVD
M32           AKSIKFPTWPPVRAFRKNIVAAQKKTSDDQTDQKASSNAITSAAFVKVSMDGAPYLRKVD
              **: :.  **************:*************************************

7710          LKLYKSYQDLSDALGKMFSSFTIGNCGSQGMKDFMNESKLIDLLNGSEYVPTYEDKDGDW
M32           LKLYKSYQDLSDALGKMFSSFTIGNCGSQGMKDFMNESKLIDLLNGSEYVPTYEDKDGDW
              ************************************************************

7710          MLVGDVPWEMFVGSCKRLRIMKGSEAIGLAPRAVEKCKNRS*
M32           MLVGDVPWEMFVGSCKRLRIMKGSEAIGLAPRAVEKCKNRS*
              *****************************************
```

Table 3-5. CLUSTAL alignment of transposable elements similar to the one in *AUX/IAA16* from the M32 population. Outgroup represents the version on chromosome 2. Donor represents the conserved version found on chromosome 4. Insertion is the M32-specific version inside of *AUX/IAA16*.

```
IAA_insertion      TGTTGGGATTAAACCCTAATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCA
M32_donor          TGTTGGGATTAAACCCTAATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCG
7710_donor         TGTTGGGATTAAACCCTAATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCG
M32_outgroup       T-----------------------------------------------------------
7710_outgroup      T-----------------------------------------------------------
                   *

IAA_insertion      TACCTGAATCCATGACGAAGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCC
M32_donor          TACCTGAATCCATGACGAAGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCC
7710_donor         TACCTGAATCCATGACGAAGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCC
M32_outgroup       ------------------------------------------CCAATTCTTTATGGCTCC
7710_outgroup      ------------------------------------------CCAATTCTTTATGGCTCC
                                                             ******************

IAA_insertion      TTAGGGTTTCTACTGATGATGGGATGTCAGTGAGAATAGGAAAACCATATGAATCGGGGA
M32_donor          TTAGGGTTTCTACTGATGATGGGATGTCAGTGAGAATAGGAAAACCATATGAATCGGGGA
7710_donor         TTAGGGTTTCTACTGATGATGGGATGTCAGTGAGAATAGGAAAACCATATGAATCGGGGA
M32_outgroup       TTAGGGTTTCTACTGATGATGGGATGTCAGTGAGAATAGGAAAACCATATAAATCGGGGA
7710_outgroup      TTAGGGTTTCTACTGATGATGGGATGTCAGTGAGAATAGGAAAACCATATAAATCGGGGA
                   ************************************************** *********

IAA_insertion      CCATAACCCTTAATGTCTATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAAT
M32_donor          CCATAACCCTTAATGTCTATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAAT
7710_donor         CCATAACCCTTAATGTCTATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAAT
M32_outgroup       CCATAACCCTTAATGTCTATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAAT
7710_outgroup      CCATAACCCTTAATGTCTATTTATATACATAGACTCCTTCCTA-TCCGCCCATCGTGAAT
                   ******************************************* ****************

IAA_insertion      AAGGAAAGGCCCATCGGTATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGAC
M32_donor          AAGGAAAGGCCCATCGGTATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGAC
7710_donor         AAGGAAAGGCCCATCGGTATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGAC
M32_outgroup       AAGGAAAGGCCCATCGGTATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGAC
7710_outgroup      AAGGAAAGGCCCATCG-TATCTAC-CAAATAAAAA-CTGATCCCACACTCTATAAAAGAC
                   **************** ******* ********** ************************

IAA_insertion      GTAAATAGGCCCAATAATAATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTAT
M32_donor          GTAAATAGGCCCAATAATAATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTAT
7710_donor         GTAAATAGGCCCAATAATAATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTAT
M32_outgroup       GTAAATAGGCCCAATAATAATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTAT
7710_outgroup      GTAAATAGGCCCAATA-TAATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTAT
                   **************** *******************************************

IAA_insertion      GATAGCACATAATACAATTATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAA
M32_donor          GATAGCACATAATACAATTATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAA
7710_donor         GATAGCACATAATACAATTATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAA
M32_outgroup       GATAGCACATAATACAATTATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAA
7710_outgroup      GATAGCACA--ATACAATTATAACTGAATTGCACACGA--TTGTATTTAGGTCCATAAAA
                   *********  **************************   ********************

IAA_insertion      TTTCCAACAGTCTCCCACTTGGACCAAATACAATTACAACGTGTGTGATACTTGATAATT
M32_donor          TTTCCAACAGTCTCCCACTTGGACCAAATACAATTACAACGTGTGTGATACTTGATAATT
7710_donor         TTTCCAACAGTCTCCCACTTGGACCAAATACAATTACAACGTGTGTGATACTTGATAATT
M32_outgroup       TTTCCAACAGTCTCCCACTTGGACCAAATACAATTACAACGTGTGTGATACTTGATAATT
7710_outgroup      TTTCCAACAGTCTCCCACTTGGACCAAATACAATTACAACGTGTGTGATACTTGATAATT
                   ************************************************************

IAA_insertion      GCATTATACTTTATAA-CCTTATGAGCTCAAAATTGCTATCAAAATCTCAACGTCTTTAA
M32_donor          GCATTATACTTTATAA-CCTTATGAGCTCAAAATTGCTATCAAAATCTCAACGTCTTTAA
```

```
7710_donor          GCATTATACTTTATAA-CCTTATGAGCTCAAAATTGCTATCAAAATCTCAACGTCTTTAA
M32_outgroup        GCATTATACTTTATAA-CCTTATGAGCTCAAAATTGCTATCAAAATCTCAACGTCTTTAA
7710_outgroup       GCATTATACTTTATAAACCTTATGAGCTCAAA-TTGCTATCAAAATCTCAACGTCTTTAA
                    **************** *************** ***************************

IAA_insertion       ACAATTCAGTCCATTAATTACATCAACACAGGATCAAAGAGATCTATGCTACATTTGCCG
M32_donor           ACAATTCAGTCCATTAATTACATCAACACAGGATCAAAGAGATCTATGCTACATTTGCCG
7710_donor          ACAATTCAGTCCATTAATTACATCAACACAGGATCAAAGAGATCTATGCTACATTTGCCG
M32_outgroup        ACAATTCAGTCCATTAATTACATCAACACAGGATCAAAGAGATCTATGCTACATTTGCCG
7710_outgroup       ACAATTCAGTCCATTAATTACC---ACACAGGATCAAAGAGATCTATGCTACATTTGCCG
                    *********************    ***********************************

IAA_insertion       TAACCAGACCCATCAATGGTCACAATGTCAACATAATTAACGACATGAATCAAGCATGGG
M32_donor           TAACCAGACCCATCAATGGTCACAATGTCAACATAATTAACGACATGAATCAAGCATGGG
7710_donor          TAACCAGACCCATCAATGGTCACAATGTCAACATAATTAACGACATGAATCAAGCATGGG
M32_outgroup        TAACCAGACCCATCAATGGTCACAATGTCAACATAATTAACGACATGAATCAAGCATGGG
7710_outgroup       TAACCAGACCCATCAATGGTCACAATGTCAACATAATTAACGACATGAATCA-GCATGGG
                    **************************************************** *******

IAA_insertion       TGTGTAGCATGGAAATTACATACAAATGTGATCCAAGTATGCCTATTTCCAACTGGTCCA
M32_donor           TGTGTAGCATGGAAATTACATACAAATGTGATCCAAGTATGCCTATTTCCAACTGGTCCA
7710_donor          TGTGTAGCATGGAAATTACATACAAATGTGATCCAAGTATGCCTATTTCCAACTGGTCCA
M32_outgroup        TGTGTAGCATGGAAATTACATACAAATGTGATCCAAGTATGCCTATTTCCAACTGGTCCA
7710_outgroup       TGTGTAGCATGGAAATTACATACAAATGTGATCCAAGTATGCCT-TTTCCAACTGGTCCA
                    ******************************************** ***************

IAA_insertion       CTGTAAACTTTGGTAAGATCAACAATATGATCTAATCGAAGAGTAAACCGAACCGAATAC
M32_donor           CTGTAAACTTTGGTAAGATCAACAATATGATCTAATCGAAGAGTAAACCGAACCGAATAC
7710_donor          CTGTAAACTTTGGTAAGATCAACAATATGATCTAATCGAAGAGTAAACCGAACCGAATAC
M32_outgroup        CTGTAAACTTTGGTAAGATCAACAATATGATCTAATCGAAGAGTAAACCGAACCGAATAC
7710_outgroup       CTGTAAACTTTGGTAAGATCAACAATATGATCTAATCGAAGAGTAA-CCGAACCGAATAC
                    ********************************************** *************

IAA_insertion       CTTATTTCTGCAGAAAATACCTTAAACCTTAATATCCGAATTAGAACTTCATCATACAAG
M32_donor           CTTATTTCTGCAGAAAATACCTTAAACCTTAATATCCGAATTAGAACTTCATCATACAAG
7710_donor          CTTATTTCTGCAGAAAATACCTTAAACCTTAATATCCGAATTAGAACTTCATCATACAAG
M32_outgroup        CTTATTTCTGCAGAAAATACCTTAAACCTTAATATCCGAATTAGAACTTCATCATACAAG
7710_outgroup       CTTATTTCTGCAGAAA-TACCTTAAACCTTAATATCCGAATTAGAACTTCATCATACAAG
                    **************** *******************************************

IAA_insertion       CATCAACATAACAAACTCCCACTGAAACTGTATATCCTTAATACCTTAACTGGCATGACA
M32_donor           CATCAACATAACAAACTCCCACTGAAACTGTATATCCTTAATACCTTAACTGGCATGACA
7710_donor          CATCAACATAACAAACTCCCACTGAAACTGTATATCCTTAATACCTTAACTGGCATGACA
M32_outgroup        CATCAACATAACAAACTCCCACTGAAACTGTATATCCTTAATACCTTAACTGGCATGACA
7710_outgroup       CATCAACATAACAAACTCCCACTGAAACTGTATATCCTTAATACCTTAACTGGCATGACA
                    ************************************************************

IAA_insertion       CACATGAATAATGTGCTCATGAAATACTTTAAGTATTTAACTTTAGTAAGCGGATCCGCA
M32_donor           CACATGAATAATGTGCTCATGAAATACTTTAAGTATTTAACTTTAGTAAGCGGATCCGCA
7710_donor          CACATGAATAATGTGCTCATGAAATACTTTAAGTATTTAACTTTAGTAAGCGGATCCGCA
M32_outgroup        CACATGAATAATGTGCTCATGAAATACTTTAAGTATTTAACTTTAGTAAGCGGATCCGCA
7710_outgroup       CACATGAATAATGTGCTCATGAAATACTTTAAGTATTTAACTTTAGTAAGCGGATCCGCA
                    ************************************************************

IAA_insertion       ATAATGGAGTTGTATTATGATGCTTTAAGTACAATAATTTTACTCTTCAAACTTTCACAA
M32_donor           ATAATGGAGTTGTATTATGATGCTTTAAGTACAATAATTTTACTCTTCAAACTTTCACAA
7710_donor          ATAATGGAGTTGTATTATGATGCTTTAAGTACAATAATTTTACTCTTCAAACTTTCACAA
M32_outgroup        ATAATGGAGTTGTATTATGATGCTTTAAGTACAATAATTTTACTCTTCAAACTTTCACAA
7710_outgroup       ATAATGGAGTTGTATTATGATGCTTTAAGTACAATAATTTTACTCTTCAAACTTTCACA-
                    *********************************************************** 

IAA_insertion       CACTAAAGCCATGATAAAAATTTAACCGTATTTGGATTTCAAGCTACCTTGACATCCACC
M32_donor           CACTAAAGCCATGATAAAAATTTAACCGTATTTGGATTTCAAGCTACCTTGACATCCACC
```

```
7710_donor          CACTAAAGCCATGATAAAAATTTAACCGTATTTGGATTTCAAGCTACCTTGACATCCACC
M32_outgroup        CACTAAAGCCATGATAAAAATTTAACCGTATTTGGATTTCAAGCTACCTTGACATCCACC
7710_outgroup       CACTAAAGCCATGATAAAAATTTAACCGTATTTGGATTTCAAGCTACCTTGACATCCACC
                    ************************************************************

IAA_insertion       AAAAACGGAGTCTAATAGCTGTAAGTTCTCTCTAATGGATCTCTTTAATGATACTTATTC
M32_donor           AAAAACGGAGTCTAATAGCTGTAAGTTCTCTCTAATGGATCTCTTTAATGATACTTATTC
7710_donor          AAAAACGGAGTCTAATAGCTGTAAGTTCTCTCTAATGGATCTCTTTAATGATACTTATTC
M32_outgroup        AAAAACGGAGTCTAATAGCTGTAAGTTCTCTCTAATGGATCTCTTTAATGATACTTATTC
7710_outgroup       AAAA-CGGAGTCTAATAGCTGTAAGTTCTCTCTAATG-ATCTCTTTAATGATACTTATTC
                    **** ******************************** **********************

IAA_insertion       TTGAGGTTGTTGAGTTTGCTCATTGGAATTTTCCTTTTGAGATCCTAAATAGGGGAATAA
M32_donor           TTGAGGTTGTTGAGTTTGCTCATTGGAATTTTCCTTTTGAGATCCTAAATAGGGGAATAA
7710_donor          TTGAGGTTGTTGAGTTTGCTCATTGGAATTTTCCTTTTGAGATCCTAAATAGGGGAATAA
M32_outgroup        TTGAGGTTGTTGAGTTTGCTCATTGGAATTTTCCTTTTGAGATCCTAAATAGGGGAATAA
7710_outgroup       TTGAGGTTGTTGAGTTTGCTCATTGGAATTTTCCTTTTGAGATCCTAAATAGGGGAATAA
                    ************************************************************

IAA_insertion       ATCCTTATCTTCCCCCAAACTCAACATCCTCAAAGGACAAGCATCTATTGTATCCAAAAA
M32_donor           ATCCTTATCTTCCCCCAAACTCAACATCCTCAAAGGACAAGCATCTATTGTATCCAAAAA
7710_donor          ATCCTTATCTTCCCCCAAACTCAACATCCTCAAAGGACAAGCATCTATTGTATCCAAAAA
M32_outgroup        ATCCTTATCTTCCCCCAAACTCAACATCCTCAAAGGACAAGCATCTATTGTATCCAAAAA
7710_outgroup       ATCCTTATCTTCCCCCAAACTCAACATCCTCAAAGGACAAGCATCTATTGTATCCAAAAA
                    ************************************************************

IAA_insertion       TTGATCCAATAGTGAGACTATAAAAAAAAATGTATCCTTTATTATTATCTTCATCAATAT
M32_donor           TTGATCCAATAGTGAGACTATAAAAAAAA-TGTATCCTTTATTATTATCTTCATCAATAT
7710_donor          TTGATCCAATAGTGAGACTATAAAAAAAA-TGTATCCTTTATTATTATCTTCATCAATAT
M32_outgroup        TTGATCCAATAGTGAGACTATAAAAAAAA-TGTATCCTTTATCATTATCTTCATCAATAT
7710_outgroup       TTGATCCAATAGTGAGACTATAAAAAAAA-TGTATCCTTTATCATTATCTTCATCAATAT
                    ***************************** ************ *****************

IAA_insertion       CTTTAGTGAAAGATGTCAAGTGAGCACTTTATATCTTTCATTGTTTCAATCTCTCTTCCT
M32_donor           CTTTAGTGAAAGATGTCAAGTGAGCACTTTATATCTTTCATTGTTTCAATCTCTCTTCCT
7710_donor          CTTTAGTGAAAGATGTCAAGTGAGCACTTTATATCTTTCATTGTTTCAATCTCTCTTCCT
M32_outgroup        CTTTAGTGAAAGATGTCAAGTGAGCACTTTATATCTTTCATTGTTTCAATCTCTCTTCCT
7710_outgroup       CTTTAGTGAAAGATGTCAAGTGAGCACTTTATATCTTTCATTGTTTCAATCTCTCTTCCT
                    ************************************************************

IAA_insertion       CCGAATCAACATGAGTAGTGAACTCATTTAGAGGTTCACTTCTCCATAACAACAATTTAG
M32_donor           CCGAATCAACATGAGTAGTGAACTCATTTAGAGGTTCACTTCTCCATAACAACAATTTAG
7710_donor          CCGAATCAACATGAGTAGTGAACTCATTTAGAGGTTCACTTCTCCATAACAACAATTTAG
M32_outgroup        CCGAATCAACATGAGTAGTGAACTCATTTAGAGGTTCACTTCTCCATAACAACAATTTAG
7710_outgroup       CCGAATCAACATGAGTAGTGAACTCATTTAGAG-TTCACTTCTCCATAACAA---TTTAG
                    ********************************* ******************   *****

IAA_insertion       GAAGAGATTTAAAAACAAATGCACAAGCAAGTCGTTAGACTCCTCCAATTTAAGTGCCCT
M32_donor           GAAGAGATTTAAAAACAAATGCACAAGCAAGTCGTTAGACTCCTCCAATTTAAGTGCCCT
7710_donor          GAAGAGATTTAAAAACAAATGCACAAGCAAGTCGTTAGACTCCTCCAATTTAAGTGCCCT
M32_outgroup        GAAGAGATTTAAAAACAAATGCACAAGCAAGTCGTTAGACTCCTCCAATTTAAGTGCCCT
7710_outgroup       GAAGAGATTTAAAA-CAAATGCACAAGCAAGTCGTTAGACTCCTCCAATTTAAGTGCCCT
                    ************** *********************************************

IAA_insertion       TAAATCATGAAACAATATCACAAATTTCCATAATGTGCTCCCTTAATTTATTTATTTTCC
M32_donor           TAAATCATGAAACAATATCACAAATTTCCATAATGTGCTCCCTTAATTTATTTATTTTCC
7710_donor          TAAATCATGAAACAATATCACAAATTTCCATAATGTGCTCCCTTAATTTATTTATTTTCC
M32_outgroup        TAAATCATGAAACAATATCACAAATTTCCATAATGTGCTCCCTTAATTTATTTATTTTCC
7710_outgroup       TAA-TCATGAAACAATATCACAAATT-CCATAATGTGCTCCCTTAATTTATTTATTTTCC
                    *** ********************** *********************************

IAA_insertion       CCTTAATACCTTTTTCCCTTATGGAATTTAAGTCCAACAAGGTTGTACTCAATTCCCCTG
M32_donor           CCTTAATACCTTTTTCCCTTATGGAATTTAAGTCCAACAAGGTTGTACTCAATTCCCCTG
```

```
7710_donor          CCTTAATACCTTTTTCCCTTATGGAATTTAAGTCCAACAAGGTTGTACTCAATTCCCCTG
M32_outgroup        CCTTAATACCTTTTTCCCTTATGGAATTTAAGTCCAACAAGGTTGTACTCAATTCCCCTA
7710_outgroup       CCTTAATACCTTTTTCCCTTATGGAATTTAAGTCCAACAAGGTTGTACTCA-TTCCCCTA
                    *************************************************** ******* 

IAA_insertion       ATCTTTATAGCAAAACACTTTAAGCTCTTTAGAGAACAATTTGACATCTTTCTCTTTCAT
M32_donor           ATCTTTATAGCAAAACACTTTAAGCTCTTTAGAGAACAATTTGACATCTTTCTCTTTCAT
7710_donor          ATCTTTATAGCAAAACACTTTAAGCTCTTTAGAGAACAATTTGACATCTTTCTCTTTCAT
M32_outgroup        ATCTTTATAGCAAAACACTTTAAGCTCTTTAGAGAACAATTTGACATCTTTCTCTTTCAT
7710_outgroup       ATCTTTATAGCAAA-CACTTTAAGCTCTTTAGAGAACAATTTGACATCTTTCTCTTTCAT
                    ************** *********************************************

IAA_insertion       TTGGACATACATAATGTCCCCTAATAACTCCGAAATTGTACTTTAAATTGCACTTTACAT
M32_donor           TTGGACATACATAATGTCCCCTAATAACTCCGAAATTGTACTTTAAATTGCACTTTACAT
7710_donor          TTGGACATACATAATGTCCCCTAATAACTCCGAAATTGTACTTTAAATTGCACTTTACAT
M32_outgroup        TTGGACATACATAATGTCCCCTAATAACTCCGAAATTGTACTTTAAATTGCACTTTACAT
7710_outgroup       TTGGACATACATAATGTCCCCTAATAACTCCGAAATTGTACTTTAAATTGCACTTTACAT
                    ************************************************************

IAA_insertion       GATCATATACGACCTATCCAATTTGAGCATTCTCACTTCTCCATATTCCCCTTTAGTAGG
M32_donor           GATCATATACGACCTATCCAATTTGAGCATTCTCACTTCTCCATATTCCCCTTTAGTAGG
7710_donor          GATCATATACGACCTATCCAATTTGAGCATTCTCACTTCTCCATATTCCCCTTTAGTAGG
M32_outgroup        GATCATATACGACCTATGCAATTTGAGCATTCTCACTTCTCCATATTCCCCTTTAGTAGG
7710_outgroup       GATCATATACG-CCTATGCAATTTGAGCATTCTC-CTTCTCCATATTCCCCTTTAGTAGG
                    *********** ***** **************** *************************

IAA_insertion       AGCACTAAAATCCATAAGAGTTGCGAGTGCTCAACTCGAAAAGCAAGGTCTGATCTAGAT
M32_donor           AGCACTAAAATCCATAAGAGTTGCGAGTGCTCAACTCGAAAAGCAAGGTCTGATCTAGAT
7710_donor          AGCACTAAAATCCATAAGAGTTGCGAGTGCTCAACTCGAAAAGCAAGGTCTGATCTGGAT
M32_outgroup        AGCACTAAAATCCGTAAGAGTTGCGAGTGCTCAACTCGAAAAGCAAGGTCTGATCTAGAT
7710_outgroup       --CACTAAAATCCGTAAGAGTTGCGAGTGCTCAACTCGAAAAGCAAGGTCTGATCTAGAT
                      *********** ****************************************** ***

IAA_insertion       CCATGCAAAAAAAAAAAAAAAA-CCAAAACCTTCAATTTATTATCCTTAATCAGATTAAT
M32_donor           CCATGCAAAAAAAAAAAAAAA--CCAAAACCTTCAATTTATTATCCTTAATCAGATTAAT
7710_donor          CCATGCAAAAAAAAAAAAAAAAACCAAAACCTTCAATTTATTATCCTTAATCAGATTAAT
M32_outgroup        CCATGCAAAAAAAAAAAAAA---CCAAAACCTTCAATTTATTATCCTTAATCAGATTAAT
7710_outgroup       CCATGCAAAAAAAAAAAA-----CCAAAACCTTCA-TTTATTATCCTTAATCAGATTAAT
                    ******************     ************ ************************

IAA_insertion       ATATTGACAATGTATCCCATCATGTACAAGAGACCTAACACAACATTAATTTTTT-AGTC
M32_donor           ATATTGACAATGTATCCCATCATGTACAAGAGACCTAACACAACATTAATTTTTT-AGTC
7710_donor          ATATTGACAATGTATCCCATCATGTACAAGAGACCTAACACAACATTAATTTTTT-AGTC
M32_outgroup        ATATTGACAATGTATCCCATCATGTACAAGAGACCTAACACAACATTAATTTTTTTAGTC
7710_outgroup       ATATTGACAATGTATCCCATCATGTACAAGAGACCTAACACAACATTAATTTTTTTAGTC
                    ******************************************************* ****

IAA_insertion       TTTGGACAGAAAAATTAACTTGTAAGTGGTCCTCTTGGTGTAGTAATCAAATACTAACAA
M32_donor           TTTGGACAGAAAAATTAACTTGTAAGTGGTCCTCTTGGTGTAGTAATCAAATACTAACAA
7710_donor          TTTGGACAGAAAAATTAACTTGTAAGTGGTCCTCTTGGTGTAGTAATCAAATACTAACAA
M32_outgroup        TTTGGACAGACAAATTAACTTGTAAGTGGTCCTCTTGGTGTAGTAATCAAATACTAACAA
7710_outgroup       TTTGGACAGACAAATTAACTTGTAAGTGGTCCTCTTGGTGTAGTAATCAAATACTAACAA
                    ********** *************************************************

IAA_insertion       TAAATTCATGTCAATGTTAAGCCTTTCTTTGGACCGACTTAACAACGCATGAAAAACCGA
M32_donor           TAAATTCATGTCAATGTTAAGCCTTTCTTTGGACCGACTTAACAACGCATGAAAAACCGA
7710_donor          TAAATTCATGTCAATGTTAAGCCTTTCTTTGGACCGACTTAACAACGCATGAAAAACCGA
M32_outgroup        TAAATTCATGTCAATGTTAAGCCTTTCTTTGGATCGACTTAACAACGCATGAAAAACCGA
7710_outgroup       TAAATTCATGTCAATGTTAAGCCTTTCTTTGGATCGACTTAAC---GCATGAAAAACCGA
                    ********************************* *********   **************

IAA_insertion       ACAATTGTTACATATTTATTACCACAAGCATGTTTAATTTCCGGCCAAATAATAGCCTTC
M32_donor           ACAATTGTTACATATTTATTACCACAAGCATGTTTAATTTCCGGCCAAATAATAGCCTTC
```

```
7710_donor          ACAATTGTTACATATTTATTACCACAAGCATGTTTAATTTCCGGCCAAATAATAGCCTTC
M32_outgroup        ACAATTGTTACATATTTATTACCACAAGCATGTTTAATTTCCGGCCAAATAATAGCCTTC
7710_outgroup       ACAATTGTTACATATTTATTACCACAAGCATGTTTAATTTCCGGCCAAATAATAGCCTTC
                    ************************************************************

IAA_insertion       CTTTGGGCCGACCATTATCCGCATGAAAACTAAACACACACAGGTATCTCAATATTTCAA
M32_donor           CTTTGGGCCGACCATTATCCGCATGAAAACTAAACACACACAGGTATCTCAATATTTCAA
7710_donor          CTTTGGGCCGACCATTATCCGCATGAAAACTAAACACACACAGGTATCTCAATATTTCAA
M32_outgroup        CTTTGGGCCGACCATTATCCGCATGAAAACTAAACACACACAGGTATCTCAATATTTCAA
7710_outgroup       CTTTGGGCCGACCATTATCCGCATGAAAACTAAACACACACAGGTATCTCAATATTTCAA
                    ************************************************************

IAA_insertion       ATTAATTGATCTACACAAAAGAGGTTACTTTGGCAACATATTGTTTCAATCAATTAATCC
M32_donor           ATTAATTGATCTACACAAAAGAGGTTACTTTGGCAACATATTGTTTCAATCAATTAATCC
7710_donor          ATTAATTGATCTACACAAAAGAGGTTACTTTGGCAACATATTGTTTCAATCAATTAATCC
M32_outgroup        ATTAATTGATCTACACAAAAGAGGTTACTTTGGCAACATATTGTTTCAATCAATTAATCC
7710_outgroup       -TTAATTGATCTACACAAAAGAG--TACTTTGGCAACATAT-GTTTCAATCAATTAATCC
                     **********************  **************** ******************

IAA_insertion       AAAATACTAGACCCAACTTAATAACCAAATTGCCAATTATTACAATCCGAAATATATGAT
M32_donor           AAAATACTAGACCCAACTTAATAACCAAATTGCCAATTATTACAATCCGAAATATATGAT
7710_donor          AAAATACTAGACCCAACTTAATAACCAAATTGCCAATTATTACAATCCGAAATATATGAT
M32_outgroup        AAAATACTAGACCCAACTTAATAACCAAATTGCCAATTATTACAATCCGAAATATATGAT
7710_outgroup       AAAATACTAGACCCAACTTAATAACCAAATTGCCAAT-ATTACAATCCGAA-TATATGA-
                    ************************************* ************* ******* 

IAA_insertion       GAATCCATAATTAGTCATGTGTTTATGAATCCATGACCTCAAATAGATAAGATGAAGCGC
M32_donor           GAATCCATAATTAGTCATGTGTTTATGAATCCATGACCTCAAATAGATAAGATGAAGCGC
7710_donor          GAATCCATAATTAGTCATGTGTTTATGAATCCATGACCTCAAATAGATAAGATGAAGCGC
M32_outgroup        GAATCCATAATTAGTCATGTGTTTATGAATCCATGACCTCAAATAGATAAGATGAAGCGC
7710_outgroup       --ATCCATAATTAGTCATGTGTTTATGAATCCATGACCTCAA-TAGATAAGATGAAGCGC
                      **************************************** *****************

IAA_insertion       AAAATAAGTTGTAAACTGTAACACTAAAAATTAATCGATGGATTCATAATTCATAGCAAC
M32_donor           AAAATAAGTTGTAAACTGTAACACTAAAAATTAATCGATGGATTCATAATTCATAGCAAC
7710_donor          AAAATAAGTTGTAAACTGTAACACTAAAAATTAATCGATGGATTCATAATTCATAGCAAC
M32_outgroup        AAAATAAGTTGTAAACTGTAACACTAAAAATTAATCGATGGATTCATAATTCATAGCAAC
7710_outgroup       AAA-TAAGTTGTAAACTGTAACACTAAAA-TTAATCGCTGGATTCATAATTCATAGCACA
                    *** ************************* ******* ********************  

IAA_insertion       ACAGCAGCGACTAACAATTGAATTAAGCAATAAATTGTTCTGTCTTATATATGGAGTAAG
M32_donor           ACAGCAGCGACTAACAATTGAATTAAGCAATAAATTGTTCTGTCTTATATATGGAGTAAG
7710_donor          ACAGCAGCGACTAACAATTGAATTAAGCAATAAATTGTTCTGTCTTATATATGGAGTAAG
M32_outgroup        ACAGCAGCGACTAACAATTGAATTAAGCAATAAATTGTTCTGTCTTATATATGGAGTAAG
7710_outgroup       CCAGCA-CGACTAACAATTGAATTAAGCAATAAATT--TCTGTCTTATATATGGAGTAAG
                     ***** *****************************  **********************

IAA_insertion       ATTAAAATACCAAAGACTTTAACC--ACCAACACCCAGGCCAGGTCAAAAAATCAACCGA
M32_donor           ATTAAAATACCAAAGACTTTAACC--ACCAACACCCAGGCCAGGTCAAAAAATCAACCGA
7710_donor          ATTAAAATACCAAAGACTTTAACC--ACCAACACCCAGGCCAGGTCAAAAAATCAACCGA
M32_outgroup        ATTAAAATACCAAAGACTTTAACC--ACCAACACCCAGGCCAGGTCAAAAAATCAACCGA
7710_outgroup       ATTAAAA--CCAAAGACTTTAACCCCACCAACACCCAGGCCAGGTCAAAAAATCAACCGA
                    *******  ***************  **********************************

IAA_insertion       AATAAATAAAGTCAAGCTACATATATCTAGTGACTTTATCGATTCTAATCTTAAGAAAGG
M32_donor           AATAAATAAAGTCAAGCTACATATATCTAGTGACTTTATCGATTCTAATCTTAAGAAAGG
7710_donor          AATAAATAAAGTCAAGCTACATATATCTAGTGACTTTATCGATTCTAATCTTAAGAAAGG
M32_outgroup        AATAAATAAAGTCAAGCTACATATATCTAGTGACTTTATCGATTCTAATCTTAAGAAAGG
7710_outgroup       AATAAATAA-GTCAAGCTACATATATCTAGTGACTTTATCGATTCTAATCTTAAGAAAGG
                    ********* **************************************************

IAA_insertion       AATCACCAAATCAATGCCAAAACTAAAAAATAATCAGAACGATGACCAACCATATTTGAA
M32_donor           AATCACCAAATCAATGCCAAAACTAAAAAATAATCAGAACGATGACCAACCATATTTGAA
```

```
7710_donor          AATCACCAAATCAATGCCAAAACTAAAAAATAATCAGAACGATGACCAACCATATTTGAA
M32_outgroup        AATCACCAAATCAATGCCAAAACTAAAAAATAATCAGAACGATGACCAACCATATTTGAA
7710_outgroup       AATCACCAAATCAATGCCAAAACTAAAAAATAATCAGAACGATGACCA-CCATATTTGAA
                    ************************************************ ***********

IAA_insertion       GCACCCACAAAGCCTCATATGTCGAAATAAAATATCTGAAGTAAACTGTAAGGGATATTC
M32_donor           GCACCCACAAAGCCTCATATGTCGAAATAAAATATCTGAAGTAAACTGTAAGGGATATTC
7710_donor          GCACCCACAAAGCCTCATATGTCGAAATAAAATATCTGAAGTAAACTGTAAGGGATATTC
M32_outgroup        GCACCCACAAAGCCTCATATGTCGAAATAAAATATCTGAAGTAAACTGTAAGGGATATTC
7710_outgroup       GCACCCACAAAGCCTCATATGTCGAAATAAAATATCTGAAGTAAACTGTAAGGGATATTC
                    ************************************************************

IAA_insertion       AATTAAAGCATAATAGGGAATCAAAACCCAAATTAAATTTGGATGCTAACAAAGCCAAAA
M32_donor           AATTAAAGCATAATAGGGAATCAAAACCCAAATTAAATTTGGATGCTAACAAAGCCAAAA
7710_donor          AATTAAAGCATAATAGGGAATCAAAACCCAAATTAAATTTGGATGCTAACAAAGCCAAAA
M32_outgroup        AATTAAAGCATAATAGGGAATCAAAACCCAAATTAAATTTGGATGCTAACAAAGCCAAAA
7710_outgroup       AATTAA-GCATAATAGGGAATCAAAACCCAAATTAAATTTGGATGCTAACAAAGCCAAAA
                    ****** *****************************************************

IAA_insertion       ACATACACACACAACTGTCAACTGGATCGAATTGTGCATGTCAAATTGTCAATCATGACA
M32_donor           ACATACACACACAACTGTCAACTGGATCGAATTGTGCATGTCAAATTGTCAATCATGACA
7710_donor          ACATACACACACAACTGTCAACTGGATCGAATTGTGCATGTCAAATTGTCAATCATGACA
M32_outgroup        ACATACACACACAACTGTCAACTGGATCGAATTGTGCATGTCAAATTGTCAATCATGACA
7710_outgroup       ACATACACACACAACTGTCAACTGGATCGA-TTGTGCATGTCAA-TTGTCAATCATGACA
                    ****************************** ************* ***************

IAA_insertion       ATTGAGAAATAATTGATAACATACATCTGCCTTTTAGGCCTCGTCAATTAACAAATACTC
M32_donor           ATTGAGAAATAATTGATAACATACATCTGCCTTTTAGGCCTCGTCAATTAACAAATACTC
7710_donor          ATTGAGAAATAATTGATAACATACATCTGCCTTTTAGGCCTCGTCAATTAACAAATACTC
M32_outgroup        ATTGAGAAATAATTGATAACATACATCTGCCTTTTAGGCCTCGTCAATTAACAAATACTC
7710_outgroup       ATTGAGAAATAATTGATAACATACATCTGCCTTTTAGGCCTCGTCAATAAC--AATACTC
                    ************************************************ *   *******

IAA_insertion       AAACTCAATAGTTGATCACAAAACAAATTCGAAGTTCTGAAAGAAAAATACTTCGTACCA
M32_donor           AAACTCAATAGTTGATCACAAAACAAATTCGAAGTTCTGAAAGAAAAATACTTCGTACCA
7710_donor          AAACTCAATAGTTGATCACAAAACAAATTCGAAGTTCTGAAAGAAAAATACTTCGTACCA
M32_outgroup        AAACTCAATAGTTGATCACAAAACAAATTCGAAGTTCTGAAAGAAAAATACTTCGTACCA
7710_outgroup       AAACTCAATAGTTGATCACAAAACAAATTCGA-GTTCTGAAAGAAAAATACTTCGTACCA
                    ******************************** ***************************

IAA_insertion       AATAATTTAATATGAAGAAGGAGATCCATAAAGAATTGGACTTGTTGGGATTAAACCCTA
M32_donor           AATAATTTAATATGAAGAAGGAGATCCATAAAGAATTGGACTTGTTGGGATTAAACCCTA
7710_donor          AATAATTTAATATGAAGAAGGAGATCCATAAAGAATTGGACTTGTTGGGATTAAACCCTA
M32_outgroup        AATAATTTAATATGAAGAAGGAGATCCATAAAGAATTGGACTAGTTGGGATTAAACCCTA
7710_outgroup       A-TAATTTAATATGAAGAAGGAGATCCATAAAGAATTGGACTAGTTGGGATTAA-CCCTA
                    * **************************************** *********** *****

IAA_insertion       ATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCATACCTGAATCCATGACGA
M32_donor           ATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCATACCTGAATCCATGACGA
7710_donor          ATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCGTACCTGAATCCATGACGA
M32_outgroup        ATTCATATCAAATTAGTCGGTAATAATAGATCTGCAGAAGCATACCTGAATCCATGACGA
7710_outgroup       ATTCATATCAAATTAGTCG-TAATAATAGATCTGCAGAAGCATACCTGAATCCATGACGA
                    ******************* ********************* ******************

IAA_insertion       AGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCCTTAGGGTTTCTACTGATG
M32_donor           AGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCCTTAGGGTTTCTACTGATG
7710_donor          AGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCCTTAGGGTTTCTACTGATG
M32_outgroup        AGATCGGCGGTGGGTTTTGATCTTCCAATTCTTTATGGCTCCTTAGGGTTTCTACTGATG
7710_outgroup       AGATCGGCGG--GGTTTTGATCTTCCAATTCTTTATGGCTCCTTAGG-TTTCTACTGATG
                    **********  *********************************** ************

IAA_insertion       ATGGGATGTCAGTGAGAATAGGAAAACCATATGAATCGGGGACCATAACCCTTAATGTCT
M32_donor           ATGGGATGTCAGTGAGAATAGGAAAACCATATGAATCGGGGACCATAACCCTTAATGTCT
```

```
7710_donor         ATGGGATGTCAGTGAGAATAGGAAAACCATATGAATCGGGGACCATAACCCTTAATGTCT
M32_outgroup       ATGGGATGTCAGTGAGAATAGGAAAACCATATAAATCGGGGACCATAACCCTTAATGTCT
7710_outgroup      --GGGATGTCAGTGAGAATAGGAAAACCATATAAATCGGGGACCATAACCCTTAATGTCT
                     ****************************** ***************************

IAA_insertion      ATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAATAAGGAAAGGCCCATCGGT
M32_donor          ATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAATAAGGAAAGGCCCATCGGT
7710_donor         ATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAATAAGGAAAGGCCCATCGGT
M32_outgroup       ATTTATATACATAGACTCCTTCCTAATCCGCCCATCGTGAATAAGGAAAGGCCCATCGGT
7710_outgroup      ATTTATATACATAGACTCCTTCCTAATCC-CCCATCGTGAATAAGGAAAGGCCCATCG-T
                   ***************************** **************************** *

IAA_insertion      ATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGACGTAAATAGGCCCAATAAT
M32_donor          ATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGACGTAAATAGGCCCAATAAT
7710_donor         ATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGACGTAAATAGGCCCAATAAT
M32_outgroup       ATCTACACAAATAAAAAACTGATCCCACACTCTATAAAAGACGTAAATAGGCCCAATAAT
7710_outgroup      ATCTACACAAATAAAAAACTGATCCCACACTCTATAAAA--CGTAAATAG-CCCAATAAT
                   ***************************************  ********* *********

IAA_insertion      AATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTATGATAGCACATAATACAAT
M32_donor          AATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTATGATAGCACATAATACAAT
7710_donor         AATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTATGATAGCACATAATACAAT
M32_outgroup       AATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTATGATAGCACATAATACAAT
7710_outgroup      AATTACTTAATTGGATCACTTTAGTTTTGGGCCACACCGTATGATAGCACATAATACAAT
                   ************************************************************

IAA_insertion      TATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAATTTCCAACATGGCC
M32_donor          TATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAATTTCCAACA-----
7710_donor         TATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAATTTCCAACA-----
M32_outgroup       TATAACTGAATTGCACACGTATTTGTATTTAGGTCCATAAAATTTCCAACA-----
7710_outgroup      TATAACTGAATTGCACACGTATT-GTATTTAGGTCCATAAAATTTCCAACA-----
                   *********************** ***************************
```

Table 3-6. KASP primer set for identification of M32 allele of IAA16. The primer that binds to the wildtype has a FAM tail (green) and the primer that binds the M32 allele has a HEX tail (orange). Thermocycler conditions are 94C for 15 min; 10 cycles of 94C for 20 seconds, 61 C for 60 seconds and decreasing 0.6 C each cycle; 40 cycles of 94C for 20 seconds, 55 C for 60 seconds, quantifying FAM and HEX fluorescence at each round at 30C.

Primer name	Sequence (5'-3')
BsIAA16_intron1F_WT_KASP	GAAGGTGACCAAGTTCATGCTCTTCAGGACACAAGTTGTAGGT
BsIAA16_intron1F_M32_KASP	GAAGGTCGGAGTCAACGGATTGCCACACCGTATGATAGCAC
BsIAA16_484R_KASP	CGCTTGTGATGGCATTGCTA

Table 3-7. Polymerase chain reaction primers used to amplify *BsIAA16* from cDNA libraries. Thermocycler conditions were 95 C for 5 minutes; 35 cycles of 95 C for 15 seconds, 60 C for 15 seconds, 72 C for 1 minute; 72 C for 5 minutes.

Primer name	Sequence (5'-3')
BsIAA16_1F	ATGTTGAGTAACGAGAGAGAC
BsIAA16_843R	TCAGCTTCTGTTCTTGCACT

Table 3-8. Polymerase chain reaction primers used to clone *BsIAA16* into pFGC5941. Thermocycler conditions were 95 C for 5 minutes; 35 cycles of 95 C for 15 seconds, 60 C for 15 seconds, 72 C for 1 minute; 72 C for 5 minutes. Tails used for In-Fusion cloning are colored in green.

Primer name	Sequence (5'-3')
BsIAA16_clone_1F	TTACCATGGGGCGCGATGTTGAGTAACGAGAGAGACAAG
BsIAA16_clone_843R	GACTCACCTAGGATCTCAGCTTCTGTTCTTGCACTTC

Table 3-9. Polymerase chain reaction primers used to amplify a section of pFGC5941 that contains the transgene of interest. Thermocycler conditions were 95 C for 5 minutes; 35 cycles of 95 C for 15 seconds, 60 C for 15 seconds, 72 C for 2 minutes; 72 C for 5 minutes. Positive samples should show an amplicon of ~400bp + the length of your transgene.

Primer name	Sequence (5’-3’)
pFGC_F	CCAACCACGTCTTCAAAGCA
pFGC_R	GGCGTCTCGCATATCTCATT

Table 3-10. Quantitative PCR primers to quantify expression of various *Arabidopsis thaliana* and *Bassia scoparia* genes. Thermocycler conditions for all primer sets are 95C for 3 minutes; 40 cycles of 95C for 15 seconds, 60C for 30 seconds, quantifying SYBR fluorescence after each round at 60C.

Primer name	Sequence (5'-3')
BsIAA16_qPCR_393R	GCGACGATGTTCTTCCTGAA
BsIAA16_qPCR_209F	GCAACCAAAACTACAGCGGA
AtCyclophilin_qPCR_R	AATCGGCAACAACCACAGGC
AtCyclophilin_qPCR_F	GTCTGATAGAGATCTCACGT
AtGH3.3_qPCR_R	CGTCATTTGGAGGATTGGTTTG
AtGH3.3_qPCR_F	GAGAGCAAGGAACTCGTGTTAT
AtIAA19_qPCR_R	CTCCGTGAAAGCTCTCTTCTTC
AtIAA19_qPCR_F	AGGACTCGGGCTTGAGATAA

Figures

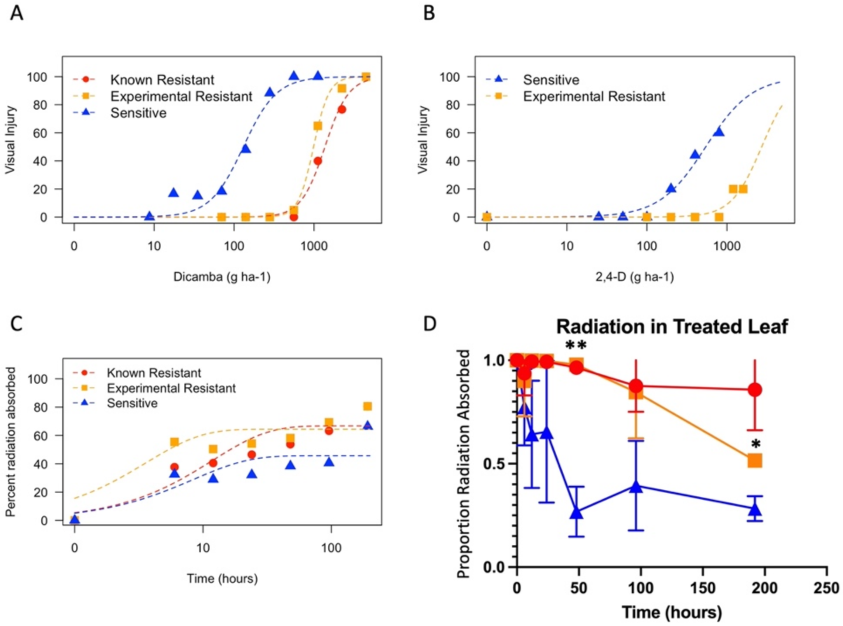


Figure 3-1. Dose response curves showing magnitude of resistance towards dicamba (A) and 2,4-D (B) for the M32 population of *Bassia scoparia* (Experimental Resistant) compared to a known dicamba-sensitive population and known dicamba-resistant population (LeClere et al., 2018; Pettinga et al., 2018). C) Time course of ^{14}C dicamba absorption showing lack of herbicide absorption is not associated with dicamba resistance in dicamba-resistant populations of *Bassia scoparia*. D) Proportion of absorbed ^{14}C dicamba retained in the treated leaf showing less dicamba exits the treated leaf in dicamba-resistant populations compared to a dicamba-sensitive population. Error bars represent standard error. *=p<0.05, **=p<0.01. Populations are represented by the same colors and symbols as in A-C.

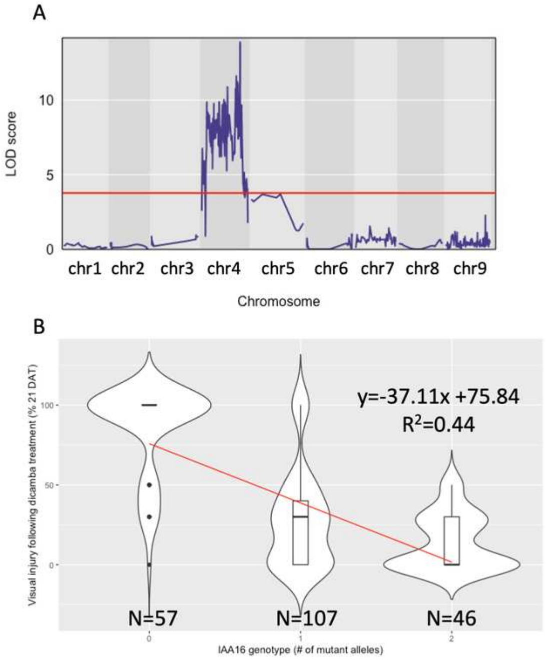


Figure 3-2. A) Quantitative trait loci (QTL) genome scan from a segregating F_3 population derived from a biparental cross of herbicide-resistant and -sensitive *Bassia scoparia* plants. Red line represents a significance threshold determined by Bonferroni-adjustment of alpha=0.05 followed by conversion to LOD as described by Nyholt (2000). B) Violin plot displaying the effect of *BsAUX/IAA16* genotype on visual injury in an independent segregating F_3 population of *Bassia scoparia*. Red line represents a linear model fitted to all points with equation listed in black. Number of observations for each genotype listed at the base of the graph.

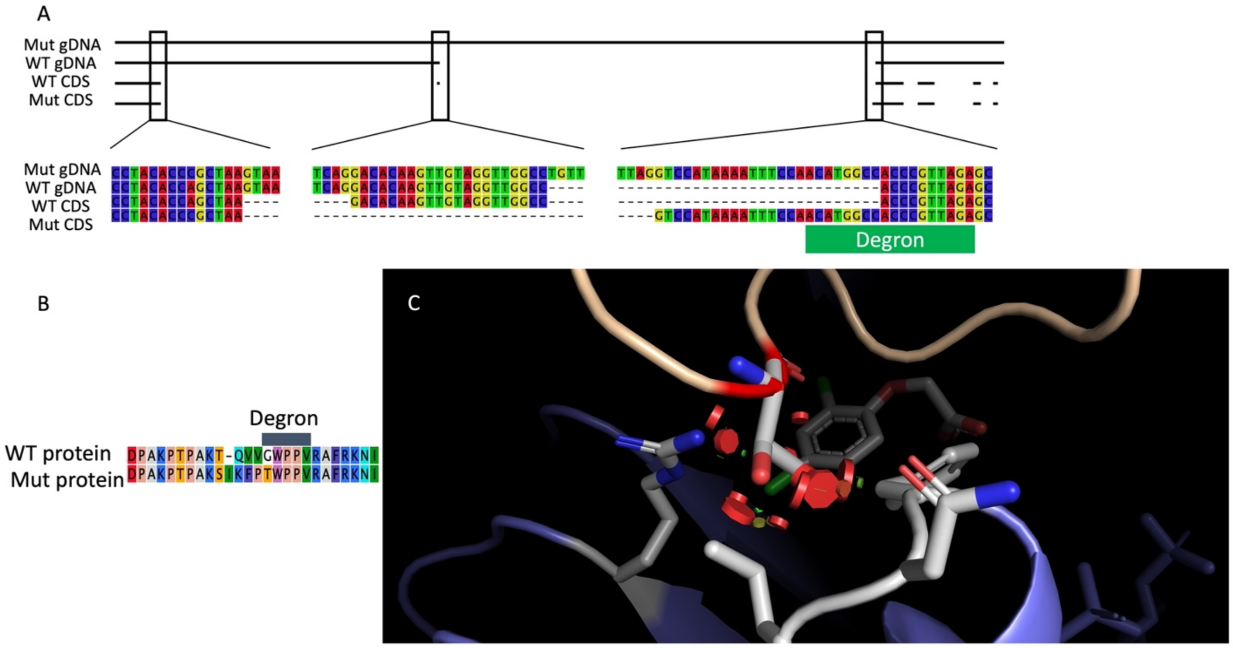


Figure 3-3. A) Genomic and coding sequence for *BsAUX/IAA16* alleles detected in a mapping population segregating for dicamba resistance. An insertion in the beginning of exon two of *BsAUX/IAA16* changes the coding sequence near the degron domain in the mutant allele. B) Amino acid sequence of mutant (Mut) and wildtype (WT) alleles of *BsAUX/IAA16* around the degron domain. C) Crystal structure of IAA7 (wheat) bound to TIR1 (slate) showing the glycine 127 residue (red) superimposed over the predicted structure if this residue is substituted for threonine, colored by element (carbon=grey, oxygen=red, nitrogen=blue, chlorine=green). Residues of TIR1 that are within 4Å of the threonine residue are colored by element, and 2,4-D is in the background, colored by element. Clashes are shown as red discs.

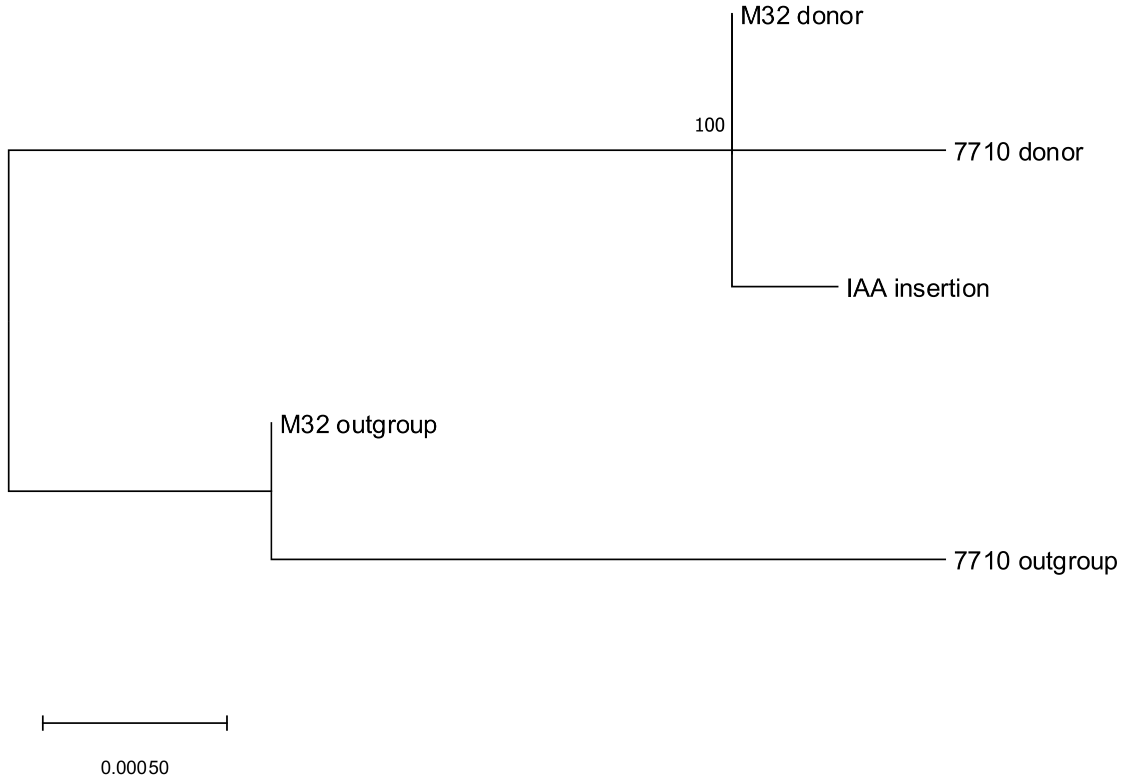


Figure 3-4. Maximum Likelihood tree showing relatedness of retrotransposons similar to the one inserted in *IAA16*. Names correspond to the descriptions in the caption of Table 3-5.

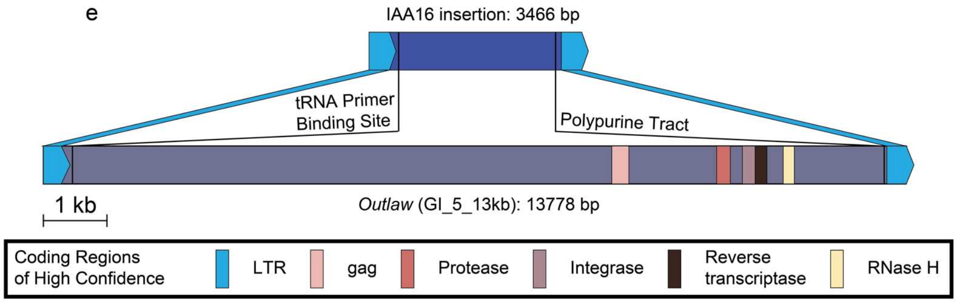


Figure 3-5. Graphical description of retrotransposon insertion from *AUX/IAA16* (top) and the autonomous version found elsewhere in the genome (bottom).

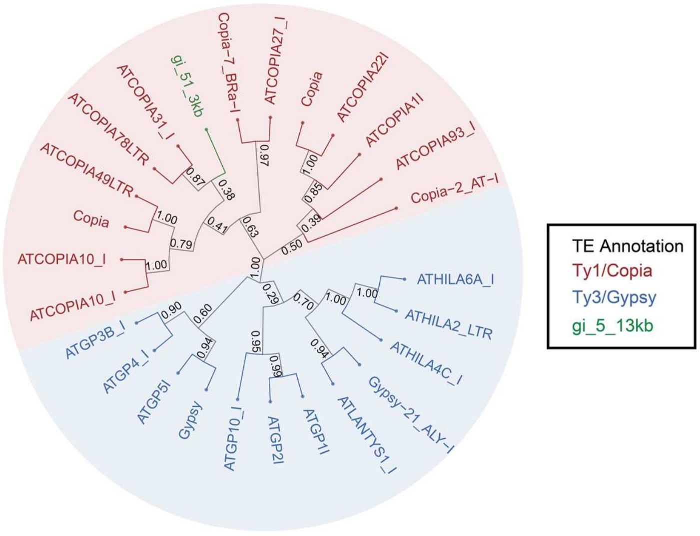


Figure 3-6. Neighbor joining tree estimating the evolutionary relatedness of long terminal repeats from gi_5_13kb and retrotransposons from *Arabidopsis thaliana*. Numbers at each branch indicate the proportion of 1000 bootstraps that support that branch.

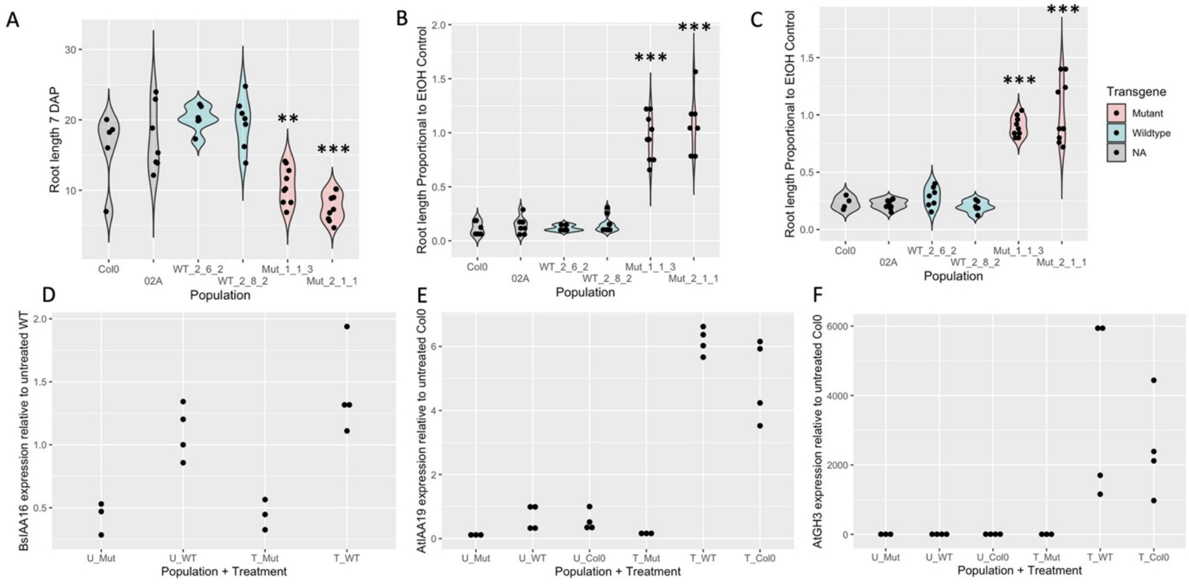


Figure 3-7. Root length of *Arabidopsis thaliana* plants expressing either *BsAUX/IAA16*$_{Mut}$ (populations starting in Mut), *BsIAA16*$_{WT}$ (populations starting in WT), or no transgene (Col0 and 02A) grown on phytoagar media. Proportional root length (compared to control) of the same populations when grown on media containing 5 µM dicamba (B) or 0.5 µM IAA (C). **=$p<0.01$, ***=$p<0.001$ Expression of the *BsAUX/IAA16* transgene (D) and the auxin response genes *AtAUX/IAA19* (E) and *AtGH3.3* (F) in plants expressing *BsAUX/IAA16*$_{Mut}$ (Mut), *BsAUX/IAA16*$_{WT}$ (WT), or no transgene (Col0) that were either treated (T) or untreated (U) with 140 g dicamba ha^{-1} six hours after treatment. Each observation is represented by a black dot.

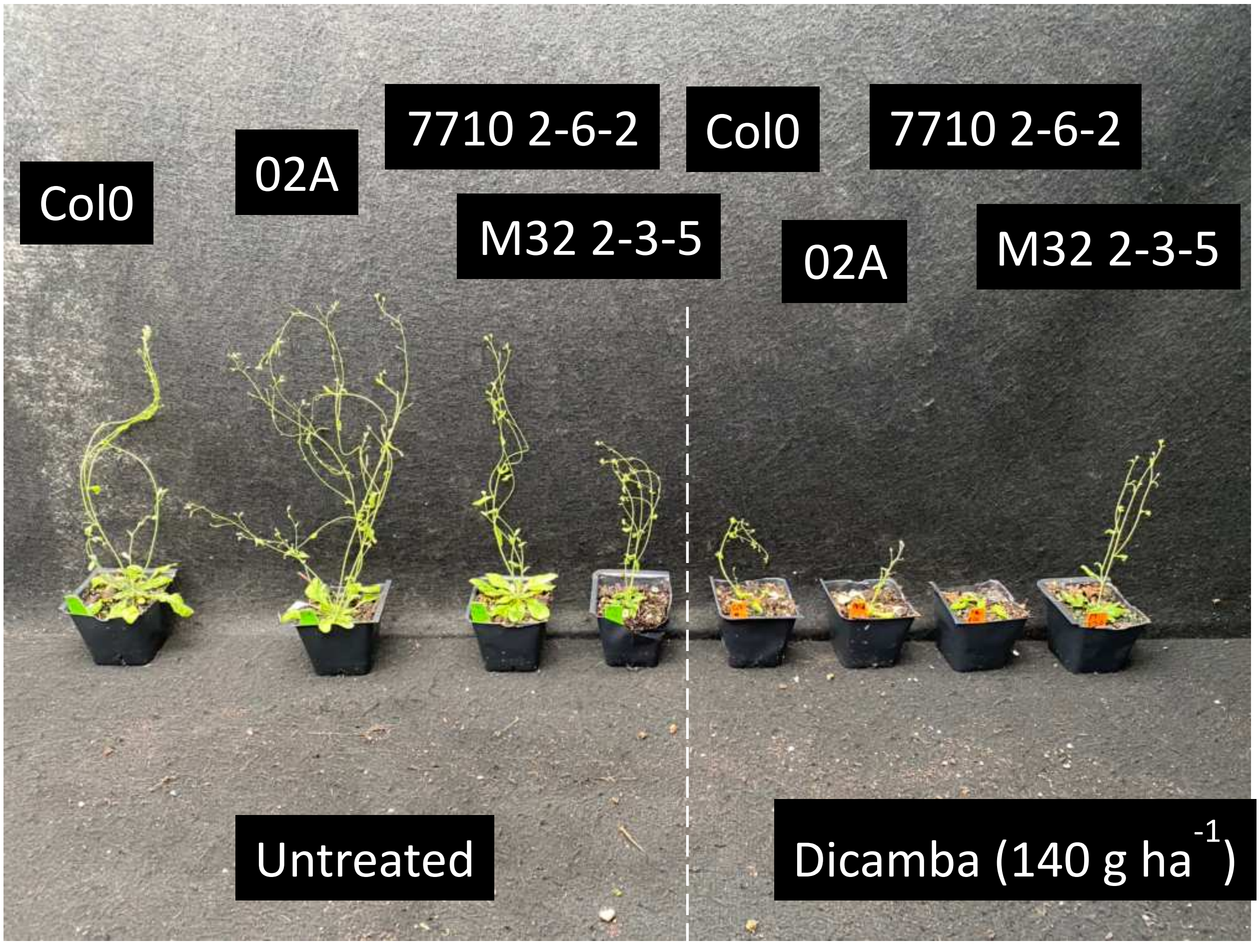


Figure 3-8. *Arabidopsis thaliana* plants (genotype Col 0) plants expressing *BsIAA16* alleles (*BsIAA16*$_{WT}$, 7710 2-6-2; *BsIAA16*$_{M32}$, M32 2-3-5) or with no transgene (02A) either untreated or treated with 140 g dicamba ha^{-1}. Photo taken 14 days after treatment.

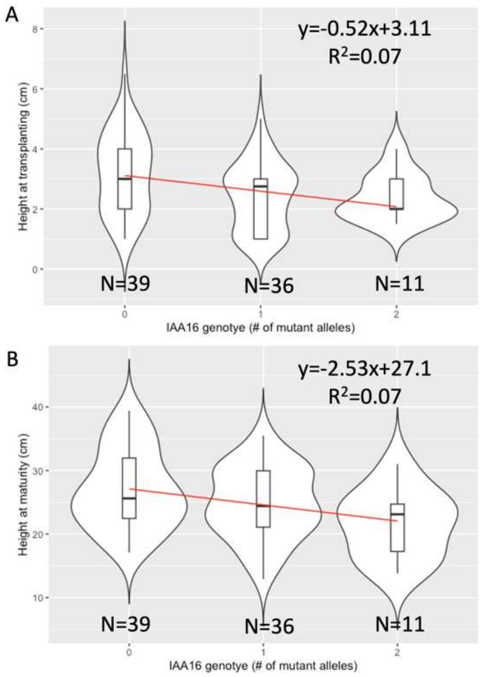


Figure 3-9. Effect of *BsAUX/IAA16* genotype on plant height either at transplanting (A) or maturity (B) of kochia plants from an F_3 population segregating for dicamba resistance.

References

Behrens, M.R., Mutlu, N., Chakraborty, S., Dumitru, R., Jiang, W.Z., Lavallee, B.J. et al. (2007) Dicamba resistance: enlarging and preserving biotechnology-based weed management strategies. *Science*, **316**, 1185-8.

Bolger, A.M., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114-20.

Broman, K.W., Gatti, D.M., Simecek, P., Furlotte, N.A., Prins, P., Sen, Ś. et al. (2019) R/qtl2: software for mapping quantitative trait loci with high-dimensional data and multiparent populations. *Genetics*, **211**, 495-502.

Busi, R., Goggin, D.E., Heap, I.M., Horak, M.J., Jugulam, M., Masters, R.A. et al. (2018) Weed resistance to synthetic auxin herbicides. *Pest Management Science*, **74**, 2265-2276.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. et al. (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H. & Li, H. (2021) Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods*, **18**, 170-175.

Christoffoleti, P., Figueiredo, M., Peres, L., Nissen, S. & Gaines, T. (2015) Auxinic herbicides, mechanisms of action, and weed resistance: A look into recent. *Scientia Agricola*, **72**, 356-362.

Clough, S.J. & Bent, A.F. (1998) Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal*, **16**, 735-43.

de Figueiredo, M.R.A., Barnes, H., Boot, C.M., de Figueiredo, A., Nissen, S.J., Dayan, F.E. et al. (2022a) Identification of a novel 2,4-D metabolic detoxification pathway in 2,4-D-resistant waterhemp (*Amaranthus tuberculatus*). *J Agric Food Chem*, **70**, 15380-15389.

de Figueiredo, M.R.A., Küpper, A., Malone, J.M., Petrovic, T., de Figueiredo, A.B.T.B., Campagnola, G. et al. (2022b) An in-frame deletion mutation in the degron tail of auxin coreceptor IAA2 confers resistance to the herbicide 2,4-D in *Sisymbrium orientale*. *Proceedings of the National Academy of Sciences*, **119**, e2105819119.

Dharmasiri, N., Dharmasiri, S. & Estelle, M. (2005) The F-box protein TIR1 is an auxin receptor. *Nature*, **435**, 441-5.

Doyle, J. & Doyle, J. (1990) A rapid total DNA preparation procedure for fresh plant tissue. *Focus*, **12**, 13-15.

Dyer, W.E. (2018) Stress-induced evolution of herbicide resistance and related pleiotropic effects. *Pest Management Science*, **74**, 1759-1768.

Espinosa, E., Bautista, R., Larrosa, R. & Plata, O. (2024) Advancements in long-read genome sequencing technologies and algorithms. *Genomics*, **116**, 110842.

Ge, X., d'Avignon, D.A., Ackerman, J.J. & Sammons, R.D. (2010) Rapid vacuolar sequestration: the horseweed glyphosate resistance mechanism. *Pest Manag Sci*, **66**, 345-8.

Ghanizadeh, H., He, L., Griffiths, A.G., Harrington, K.C., Carbone, V., Wu, H. et al. (2024) A novel mutation in *IAA16* is associated with dicamba resistance in *Chenopodium album*. *Pest Manag Sci*, 10.1002/ps.8071

Gleason, C., Foley, R.C. & Singh, K.B. (2011) Mutant analysis in Arabidopsis provides insight into the molecular mode of action of the auxinic herbicide dicamba. *PLOS ONE*, **6**, e17245.

Gray, W.M., Kepinski, S., Rouse, D., Leyser, O. & Estelle, M. (2001) Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature*, **414**, 271-6.

Grossmann, K. (2010) Auxin herbicides: current status of mechanism and mode of action. *Pest Manag Sci*, **66**, 113-20.

Hall, N., Chen, J., Matzrafi, M., Saski, C.A., Westra, P., Gaines, T.A. et al. (2023) *FHY3/FAR1* transposable elements generate adaptive genetic variation in the *Bassia scoparia* genome. *bioRxiv*, doi:10.1101/2023.05.26.542497.

Heap, I., The International Herbicide-Resistant Weed Database. www.weedscience.org [accessed 20 June 2024].

Holmes, K.H., Lindquist, J.L., Rebarber, R., Werle, R., Yerka, M. & Tenhumberg, B. (2022) Modeling the evolution of herbicide resistance in weed species with a complex life cycle. *Ecological Applications*, **32**, e02473.

Kepinski, S. & Leyser, O. (2005) The Arabidopsis F-box protein TIR1 is an auxin receptor. *Nature*, **435**, 446-451.

Kniss, A.R., Vassios, J.D., Nissen, S.J. & Ritz, C. (2011) Nonlinear regression analysis of herbicide absorption studies. *Weed Science*, **59**, 601-610.

LeClere, S., Wu, C., Westra, P. & Sammons, R.D. (2018) Cross-resistance to dicamba, 2,4-D, and fluroxypyr in *Kochia scoparia* is endowed by a mutation in an *AUX/IAA* gene. *Proceedings of the National Academy of Sciences*, **115**, E2911-E2920.

Li, H. & Durbin, R. (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**, 1754-1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078-2079.

Ma, Q., Grones, P. & Robert, S. (2017) Auxin signaling: a big question to be addressed by small molecules. *Journal of Experimental Botany*, **69**, 313-328.

Martin, S.L., Benedict, L., Wei, W., Sauder, C.A., Beckie, H.J. & Hall, L.M. (2020) High gene flow maintains genetic diversity following selection for high *EPSPS* copy number in the weed kochia (Amaranthaceae). *Scientific Reports*, **10**, 18864.

McCauley, C.L., McAdam, S.A.M., Bhide, K., Thimmapuram, J., Banks, J.A. & Young, B.G. (2020) Transcriptomics in Erigeron canadensis reveals rapid photosynthetic and hormonal responses to auxin herbicide application. *Journal of Experimental Botany*, **71**, 3701-3709.

Montgomery, J., Morran, S., MacGregor, D.R., McElroy, J.S., Neve, P., Neto, C. et al. (2024) Current status of community resources and priorities for weed genomics research. *Genome Biology*, **25**, 139.

Nyholt, D.R. (2000) All LODs are not created equal. *Am J Hum Genet*, **67**, 282-8.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G. et al. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92-4.

Patterson, E.L., Saski, C.A., Sloan, D.B., Tranel, P.J., Westra, P. & Gaines, T.A. (2019) The draft genome of *Kochia scoparia* and the mechanism of glyphosate resistance via transposon-mediated *EPSPS* tandem gene duplication. *Genome Biol Evol*, **11**, 2927-2940.

Pettinga, D.J., Ou, J., Patterson, E.L., Jugulam, M., Westra, P. & Gaines, T.A. (2018) Increased *chalcone synthase* (*CHS*) expression is associated with dicamba resistance in *Kochia scoparia*. *Pest Management Science*, **74**, 2306-2315.

Poplin, R., Ruano-Rubio, V., DePristo, M.A., Fennell, T.J., Carneiro, M.O., Van der Auwera, G.A. et al. (2018) Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv*, 10.1101/201178201178.

Preston, C., Belles, D.S., Westra, P.H., Nissen, S.J. & Ward, S.M. (2009) Inheritance of resistance to the auxinic herbicide dicamba in kochia (*Kochia scoparia*). *Weed Science*, **57**, 43-47.

Quastel, J.H. (1950) 2,4-Dichlorophenoxyacetic acid (2,4-D) as a selective herbicide. In *Agricultural Control Chemicals* eds). American Chemical Society, pp. 244-249.

R-Core-Team, R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Ramans-Harborough, S., Kalverda, A.P., Manfield, I.W., Thompson, G.S., Kieffer, M., Uzunova, V. et al. (2023) Intrinsic disorder and conformational coexistence in auxin coreceptors. *Proc Natl Acad Sci U S A*, **120**, e2221286120.

Ravet, K., Sparks, C.D., Dixon, A.L., Küpper, A., Westra, E.P., Pettinga, D.J. et al. (2021) Genomic-based epidemiology reveals independent origins and gene flow of glyphosate resistance in Bassia scoparia populations across North America. *Mol Ecol*, **30**, 5343-5359.

Ritz, C., Baty, F., Streibig, J.C. & Gerhard, D. (2015) Dose-response analysis using R. *PLOS ONE*, **10**, e0146021.

Roux, F. & Reboud, X. (2005) Is the cost of herbicide resistance expressed in the breakdown of the relationships between characters? A case study using synthetic-auxin-resistant Arabidopsis thaliana mutants. *Genet Res*, **85**, 101-10.

Ruegger, M., Dewey, E., Gray, W.M., Hobbie, L., Turner, J. & Estelle, M. (1998) The TIR1 protein of Arabidopsis functions in auxin response and is related to human SKP2 and yeast grr1p. *Genes Dev*, **12**, 198-207.

Song, Y. (2014) Insight into the mode of action of 2,4-dichlorophenoxyacetic acid (2,4-D) as an herbicide. *Journal of Integrative Plant Biology*, **56**, 106-113.

Tan, X., Calderon-Villalobos, L.I., Sharon, M., Zheng, C., Robinson, C.V., Estelle, M. et al. (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature*, **446**, 640-5.

Tatematsu, K., Kumagai, S., Muto, H., Sato, A., Watahiki, M.K., Harper, R.M. et al. (2004) MASSUGU2 encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in *Arabidopsis thaliana*. *Plant Cell*, **16**, 379-93.

Todd, O.E., Figueiredo, M.R.A., Morran, S., Soni, N., Preston, C., Kubeš, M.F. et al. (2020) Synthetic auxin herbicides: finding the lock and key to weed resistance. *Plant Sci*, **300**, 110631.

Todd, O.E., Patterson, E.L., Westra, E.P., Nissen, S.J., Araujo, A.L.S., Kramer, W.B. et al. (2024) Enhanced metabolic detoxification is associated with fluroxypyr resistance in Bassia scoparia. *Plant Direct*, **8**, e560.

Uzunova, V.V., Quareshy, M., Del Genio, C.I. & Napier, R.M. (2016) Tomographic docking suggests the mechanism of auxin receptor TIR1 selectivity. *Open Biol*, **6**, 160139.

Van der Auwera, G. & O'Connor, B. (2020) *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra (1st Edition)*. O'Reilly Media

Walsh, T.A., Neal, R., Merlo, A.O., Honma, M., Hicks, G.R., Wolff, K. et al. (2006) Mutations in an auxin receptor homolog *AFB5* and in *SGT1b* confer resistance to synthetic picolinate auxins and not to 2,4-dichlorophenoxyacetic acid or indole-3-acetic acid in Arabidopsis. *Plant Physiol*, **142**, 542-52.

Westra, E.P., Nissen, S.J., Getts, T.J., Westra, P. & Gaines, T.A. (2019) Survey reveals frequency of multiple resistance to glyphosate and dicamba in kochia (*Bassia scoparia*). *Weed Technology*, **33**, 664-672.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B. et al. (2007) A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, **8**, 973-982.

Wickham, H. (2016) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York

Wright, T.R., Shan, G., Walsh, T.A., Lira, J.M., Cui, C., Song, P. et al. (2010) Robust crop resistance to broadleaf and grass herbicides provided by *aryloxyalkanoate dioxygenase* transgenes. *Proc Natl Acad Sci U S A*, **107**, 20240-5.

Wu, C., LeClere, S., Liu, K., Paciorek, M., Perez-Jones, A., Westra, P. et al. (2021) A dicamba resistance-endowing IAA16 mutation leads to significant vegetative growth defects and impaired competitiveness in kochia (*Bassia scoparia*). *Pest Manag Sci*, **77**, 795-804.

Zenser, N., Ellsmore, A., Leasure, C. & Callis, J. (2001) Auxin Modulates the Degradation Rate of Aux/IAA Proteins. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 11795-11800.

Zhang, R.-G., Li, G.-Y., Wang, X.-L., Dainat, J., Wang, Z.-X., Ou, S. et al. (2022) TEsorter: An accurate and fast method to classify LTR-retrotransposons in plant genomes. *Horticulture Research*, **9**, uhac017.

CHAPTER III - KNOCKING OUT THE GENE *UTP-GLUCOSE-1-PHOSPHATE URIDYLYLTRANSFERASE 3* (*UGP3*) CAUSES REDUCED SULFOLIPID PRODUCTION AND OXYFLUORFEN RESISTANCE IN WHEAT

Summary

Wheat is the largest source of calories consumed by humans, worldwide. Uncontrolled weeds reduce yield in wheat production systems and threaten food security. New weed control technologies are not being discovered as rapidly as the rate of herbicide resistance evolution in weed populations. Recently, it was found that eliminating sulfolipid production in rice leads to oxyfluorfen resistance. The current study seeks to produce mutants of wheat that are resistant to the herbicide oxyfluorfen through reduced sulfolipid production. Mutant wheat lines were developed with either nonsense or frameshift mutations in all copies of the gene *UTP--glucose-1-phosphate uridylyltransferase 3* (*UGP3*) using TILLING or CRISPR-based gene editing, respectively. Using LC/MS methods, we determine that these mutants are deficient in sulfolipids, suggesting there is no secondary pathway for sulfolipid synthesis in wheat. Dose response experiments show resistance to oxyfluorfen, but not lactofen, in mutant lines. Further, LC/MS methods showed a reduction in protoporphyrin accumulation following oxyfluorfen application in the mutant lines compared to the wildtype. Molecular markers are reported for the identification of TILLING-derived mutations in *UGP3* to be used by wheat breeding programs to introgress this non-transgenic trait into locally adapted cultivars. The *UGP3* mutant lines generated and tested in this study introduce a new herbicide resistance trait in wheat. The use of oxyfluorfen in wheat complements the few other herbicide resistance traits available in wheat. If

stewarded correctly, this herbicide resistance trait will improve weed control in wheat production systems and extend the life of currently efficacious herbicides.

Introduction

The domestication and increase in productivity of bread wheat played a large role in the evolution of human civilization over time. Today, wheat products account for the largest source of calories consumed by humans, constituting 19% of total energy intake (Tadesse et al., 2019). One important advancement that allowed wheat to be so successful cultivated as a staple crop is the availability of herbicides that control weeds in wheat production systems. Most weed control in wheat systems relies on herbicides that are naturally selective. However, two herbicide resistance traits have been developed through EMS mutagenesis approaches to offer producers more management options. These include the Clearfield (Newhouse et al., 1992) and AXigen (Ostlie et al., 2015) traits, that confer resistance to imidazolinone herbicides and quizalofop-p-ethyl respectively. Glyphosate-resistant wheat has been developed through the expression of a glyphosate-insensitive form of *5-enolpyruvylshikimate 3-phosphate synthase* (Zhou et al., 1995) but has never been marketed. Over time, weeds have evolved resistance to many commonly used herbicides including ones used in Clearfield- and AXigen-traited wheat production systems (Vencill et al., 2012; Busi et al., 2013; Soni et al., 2022). The lack of discovery of herbicides with new mechanisms of action (Duke & Dayan, 2022) along with the lack of new herbicide resistance traits in wheat leaves producers with limited options for controlling herbicide resistant weed populations.

Herbicides that inhibit protoporphyrinogen oxidase (PPO) show broad spectrum control of many weed species and have been successfully used in cropping systems with natural selectivity for decades (Hao et al., 2011). There has been a recent influx of research towards the discovery of new PPO-inhibiting herbicides driven by the evolution of herbicide resistant weed populations, especially the evolution of weeds resistant to glyphosate (Barker et al., 2023). In fact, transgenic crops with robust resistance to PPO-inhibiting herbicides are currently under development and are expected to be commercialized in the coming decade (Larue et al., 2020). While PPO-inhibitor resistance evolution has been documented for 17 unique weed species (Heap, 2024), resistance has been relatively slow to evolve to these herbicides and is not as widespread compared to other herbicide modes of action. Resistance to PPO-inhibitors in weeds is usually attributed to alterations of its target sites (Patzoldt et al., 2006; Rousonelos et al., 2012; Rangani et al., 2019; Bi et al., 2020; Nie et al., 2023), but overexpression of its target site (Lermontova & Grimm, 2000) and herbicide metabolism (Dayan et al., 1997) have also been shown to be effective in conferring resistance in crops. Recently, an unexpected resistance mechanism was discovered through mutagenesis in rice. In this case, knocking out genes involved in the production of sulfoquinovosyldiacylglycerol (SQDG) resulted in resistance to the herbicide oxyfluorfen (Mckenzie et al., 2019).

Here, we recreate the SQDG-deficient phenotype in hexaploid bread wheat through gene editing and mutagenesis of the gene *UTP--glucose-1-phosphate uridylyltransferase 3* (*UGP3*). We use LC/MS methods to confirm that the mutations we make to *UGP3* are sufficient to drastically reduce SQDG concentrations in the plant. We go on to show resistance to oxyfluorfen in the mutant lines. This resistance is accompanied by a reduction in protoporphyrin accumulation following oxyfluorfen treatment. We supplement these findings with discussion about this new mechanism of herbicide resistance and considerations for crop varieties developed with this resistance trait. Future integration of this trait into elite germplasm will expand the list of weed control methods available to wheat producers.

Results and Discussion

Development of UGP3 knockout lines through gene editing and TILLING

To reduce the production of sulfolipids in wheat plants, two approaches were taken that aimed to knockout all the homoeologs of UGP3 (Figure 4-1). First, CRISPR-Cas9 transgene constructs similar to the ones described by Debernardi et al. (2020) were used to generate frameshift mutations through non-homologous end joining (NHEJ). Three different guides were used individually, but only one guide generated edits. These guides targeted regions that were conserved across all three homoeologs and towards the beginning of the gene. This allowed us to edit all three copies of *UGP3* simultaneously and ensured that crucial domains for protein function were disrupted by frameshift mutations that were generated. Edits that were fixed in the mutant line are displayed in Figure 4-2. These edits result in small insertions or deletions of one or two base pairs, shifting the reading frame of the gene for all codons downstream of these edits. Screening hundreds of T_2 plants for segregation of the CRISPR-Cas9 transgene cassette yielded no transgene-free plants. Thus, we expect many insertion sites of our transgene. If this edited line were to be used for commercial wheat variety development, a more extensive screen of segregating populations derived from backcrosses will be necessary to remove the transgenes.

In parallel, we used a crossing approach with three mutant spring wheat lines from a mutagenesis library generated and exome-capture sequenced by Krasileva et al. (2017). These three lines each have a premature stop codon introduced in the *UGP3* locus of one genome (Figure 4-1). Molecular markers for these nonsense mutations allowed us to identify hybrids and to generate a fixed line with truncated versions of all six alleles of *UGP3*. These markers (detailed in Table 4-1) will assist breeders in introgressing these alleles into elite germplasm. Though this introgression is underway, there are still many unrelated mutations from the EMS mutagenesis that are still segregating in our mutant line.

Using both mutagenesis approaches allows us to robustly examine the effects of *UGP3* knockouts and the subsequent reduction in sulfolipids in wheat. They also offer breeders options in how they can breed new wheat varieties with reduced sulfolipids. This is important given the dynamic landscape around gene edited crop regulation worldwide (Buchholzer & Frommer, 2023). Recreating these mutations and recovering a transgene-free wheat plant would, in theory, take two to three generations using a gene editing approach. This approach allows for rapid introgression of the trait into pre-adapted cultivars, especially with recent advancements in transformation efficiency for new genotypes (Debernardi et al., 2020; Wang et al., 2022; Johnson et al., 2023). Comparatively, introgressing the mutagenesis-derived alleles of *UGP3* into a new genotype would require several generations of backcrossing and then reestablishment of a line fixed at all three loci, totaling at least five or six generations. Even still, deleterious alleles tightly linked to the *UGP3* locus will likely persist in cultivars containing the mutagenesis-derived alleles.

UGP3 knockouts lead to decreased sulfolipid concentrations

To understand if these mutations to *UGP3* are sufficient to cause reduction in sulfolipid production, mutant and wildtype lines from gene editing and mutagenesis approaches were subjected to LC/MS quantification of SQDG. Both approaches resulted in near-complete elimination of SQDG (Table 4-2) with 119-fold (p=0.0012) and 21-fold (p=0.0039) reductions in

SQDG concentration in the mutant vs. wildtype lines for the mutagenesis and gene editing experiments, respectively. This result confirms that there is no other ortholog of *UGP3* in wheat or secondary sulfolipid production pathway that replaces the function of *UGP3*.

UGP3 knockout lines resist oxyfluorfen

To understand if knocking out *UGP3*, and therefore sulfolipid production, causes resistance to oxyfluorfen in wheat as it does in rice (Mckenzie et al., 2019), mutant and wildtype lines were subjected to dose response experiments. We opted for a seed soak herbicide application method because of oxyfluorfen's use as a preemergent herbicide and the limitations of soil-based resistance assays (Kerr et al., 2023). The gene edited line had a resistance factor of 5.66 ($p= 5.3x10^{-09}$) for oxyfluorfen and 1.79 ($p= 0.00026$) for lactofen (Figure 4-3; Figure 4-4; Table 4-3). While we did determine a modest reduction in lactofen sensitivity associated with the mutant line, this level of resistance may not lead to robust crop safety in the field. Importantly, the gene edited line shows no apparent difference in physiology compared to the unedited line. This suggests the lack of a fitness cost associated with reduction in sulfolipid content and oxyfluorfen resistance, similar to observations in rice. Moving forward, examination of sulfolipid knockout lines with diverse pedigrees in diverse climates will help elucidate environmental or genetic factors that affect this herbicide resistance trait in wheat. For instance, SQDG plays a role in maintaining a negatively charged lipid–water interface and stabilizing photosystem II (Yu et al., 2002; Nakajima et al., 2018). Under conditions of high temperature (Nakajima et al., 2018) or limited availability of inorganic phosphate (Yu et al., 2002), SQDG knockout lines of Arabidopsis and cyanobacteria have shown reduced growth compared to wildtype lines.

Understanding these factors will help inform placement of this herbicide resistance trait across diverse climates and guide management decisions.

Quantification of protoporphyrin accumulation following herbicide treatment

Application of herbicides that inhibit protoporphyrinogen oxidase generally result in the accumulation of protoporphyrin in sensitive plants. To understand if reduction of sulfolipid content limits this accumulation, *UGP3* wildtype and mutant lines were treated with either lactofen or oxyfluorfen and protoporphyrin was quantified. Both lines accumulated protoporphyrin after treatment with either lactofen or oxyfluorfen; however, the mutant line accumulated significantly less protoprophyrin ($p = 0.014$) compared with wildtype line when treated with oxyfluorfen (Figure 4-5). There was no significant difference in protoporphyrin levels across lines following application of lactofen. These results suggest elimination of sulfolipids interrupts herbicide action before protoporphyrin accumulation.

Our results show that elimination of sulfolipids provides resistance to oxyfluorfen in wheat, similar to rice (Mckenzie et al., 2019). We achieve this phenotype through two reverse genetics approaches including CRISPR-Cas9 mediated gene editing and EMS-based TILLING. While we did detect a modest effect towards lactofen sensitivity, this resistance mechanism seems to be specific to only oxyfluorfen. The reduction of SQDG interrupts oxyfluorfen action at or before the step of protoporphyrin accumulation as evidenced by reduced protoporphyrin accumulation in the mutant lines following oxyfluorfen application. This new resistance trait offers new weed control options to wheat producers in the face of widespread and ongoing evolution of herbicide resistant weed populations.

Materials and Methods

Plant growth conditions

Unless otherwise noted, plants in this study were grown in greenhouses at Colorado State University in 3.8-cm by 3.8-cm by 5.8-cm pots containing fine-grade potting mix (Fafard #2-SV; American Clay Works, Denver, CO). Temperatures were held between 22 C and 24 C, and supplemental light was provided by liquid halogen grow lights to ensure a photoperiod of 14h/10h. Plants were sub-irrigated every two days enough to keep the soil moist.

Design of guide RNAs

Coding and genomic sequences for genes encoding *UGP3* in the *Triticum aestivum* Refseqv2 assembly (TraesCS1A03G0736800.1, TraesCS1B03G0840600, and TraesCS1D03G0702900LC.1) were extracted from the EnsemblPlants database (Zhu et al., 2021). Sequences were aligned using CLC Sequence Viewer 8 (QIAGEN, Aarhus, Denmark) to identify sequence variance across homoeologs and splice sites. The coding sequence of the homoeolog found on chromosome 1A was submitted to CRISPR-P to identify potential guide sequences (Liu et al., 2017). The three guide sequences with the highest predicted scores that were conserved across homoeologs and unaffected by splicing were selected (sequences listed in Table 4-4) and synthesized by Integrated DNA Technologies (Coralville, Iowa, USA).

Construction and transformation of binary vectors

This process was carried out for each guide individually to produce vectors expression only one sgRNA each. Sense and antisense guide sequences were annealed by adding 3 µl of each oligo (100uM) to 24 ul water, heating to 95 C for 3 min and incubating at room temperature for

20 min. The annealed oligos (5ul) were ligated into 50 ng of plasmid JD633 (Debernardi et al., 2020) that had been digested with *Aar*I using the T4 DNA ligase protocol from New England Biolabs (Ipswich, MA, USA). Ligation product was transformed into One Shot™ TOP10 Chemically Competent *E. coli* following manufacturer specifications (ThermoFisher Scientific, Waltham, MA, USA). Individual colonies were validated using colony PCR and Sanger sequencing (Azenta; Burlington, MA, USA; primer sequences and thermocycler conditions in Table 4-4). Colonies with correctly assembled plasmids were grown in liquid LB media overnight and plasmid DNA extracted using the ZymoPURE Plasmid Miniprep Kit (Zymo Research, Irvine, CA, USA). Plasmids were again confirmed for correct construction through Sanger sequencing. Plasmids for each guide were individually transformed into chemically competent *Agrobacterium tumefaciens* (strain GV3101) following methods described by Weigel & Glazebrook (2002). Wheat callus from the Bobwhite spring wheat cultivar was transformed and regenerated following methods described by Debernardi et al. (2020). Edits in regenerated plants were identified through Sanger sequencing (primer sequences and thermocycler conditions listed in Table 4-5). Frameshift edits were fixed in T2 plants and seed was bulked to the T4 generation to generate enough seed for functional testing.

Generation of UGP3 knockout lines through TILLING

Wheat lines with premature stop codons introduced in different homoeologs of *UGP3* were identified in the Cadenza TILLING library (Krasileva et al., 2017) by BLASTing the genomic sequence of the 1A *UGP3* homoeolog using the wheat-tilling.com legacy website. These lines (1A: Cadenza1254, 1B: Cadenza1443, 1D: Cadenza 1136) were ordered from the SeedStor website. Newly developed KASP assays that detect the nonsense mutations in each homoeolog of *UGP3* (Table 4-6) were used to identify hybrids between 1A and 1B mutant lines and then hybrids between progeny of this cross and a 1D mutant. A segregating F2 population produced from this second cross was used to identify a plant homozygous for knockouts of each *UGP3* homoeolog. This plant was self-pollinated to bulk seed of the full mutant genotype.

Quantification of SQDG in mutant and wildtype plants

To determine the effect of *UGP3* knockouts on SQDG concentrations in plant tissue, seedlings of *UGP3* mutant and wildtype lines were grown. When seedlings had four leaves, above ground tissue from four plants of each population were harvested and flash frozen in liquid nitrogen. Tissue was stored at -80°C until analysis. Samples were individually ground with a mortar and pestle using liquid nitrogen to keep tissue frozen. Ground tissue (20 +/- 0.5mg) was weighed into a clean, glass 2 mL autosampler vial and capped. Chilled, -20°C 6:3:1 MTBE/methanol/water extraction solvent was then added. The samples were then briefly hand vortexed and then sonicated in a cold bath for 10 minutes. They were again hand vortexed and shaken in a cold cabinet at 4°C for 30 minutes and then centrifuged for 15 min at 2,000×g at 4°C. Supernatant (800 µL) was recovered and transferred to a clean, glass 2 mL autosampler vial. To remove all remaining tissue debris, the sample was again centrifuged and 600 µL of supernatant was collected and stored at -80°C.

A standard curve was prepared by diluting a suspended stock solution (62.5 µg/mL) of SQDG (Avanti Polar Lipids 840525P) in 100% methanol for the following curve concentrations: 6.25, 3.125, 1.5625, 0.78125, 0.390625, and 0 µg/mL. One microliter of extract was injected onto a Waters Acquity UPLC system in randomized order. Separation was achieved using a Waters ACQUITY UPLC Premier BEH Amide 1.7µm Column (1.7 µM, 2.1 x 50 mm), using a gradient from solvent A (10 mM ammonium formate in water, 0.1% formic acid) to solvent B (acetonitrile, 0.1% formic acid). Injections were made in 99% B at a flow rate of 0.8 mL/min, held at 99% B for 0.5 minutes at 0.8 mL/min, ramped to 60% A over 1.5 minutes at a flow rate of 0.8 mL/min, and then returned to 99% B over 0.5 minutes at 1 mL/min, held at 99% B for 2 minutes at 1 mL/min, and then returned to starting conditions and allowed to re-equilibrate for 0.5 minutes. The column and samples were held at 30 °C and 6 °C, respectively. The column eluent was infused into a Waters Xevo G2-XS Q-TOF-MS with an electrospray source in positive mode, scanning 50-1200 m/z at 0.1 seconds per scan, alternating between MS (6 V collision energy) and MSE mode (15-30 V ramp). Calibration was performed using sodium formate with 1 ppm mass accuracy. The capillary voltage was held at 700 V. The source temperature was held at 150 °C and the nitrogen desolvation temperature at 600 °C with a flow rate of 1000 L/hr. Means of the mutant populations were compared to the wildtype for each experiment using t-tests.

Dose response experiments

To quantify the effect of *UGP3* knockouts on sensitivity to oxyfluorfen, mutant and wildtype lines were used in dose response experiments. Ten seeds from each population were isolated in their own tea strainer (one strainer per population per treatment). The seeds were soaked in herbicide solution at differing concentrations for 24 h. Herbicide solutions used distilled water and the ALB2023 formulation of oxyfluorfen (4 lb oxyfluorfen gal^{-1}; Albaugh LLC, Ankeny, IA, USA) or the Cobra formulation of lactofen (Valent USA, San Ramon, CA, USA). After 24 hours, strainers were removed from herbicide solution and rinsed thoroughly with distilled water. Seeds were planted in trays (25 x 50 x 6.5 cm) filled with pre-soaked potting soil in furrows ~0.5 cm deep. Seeds were dusted with soil through a sieve just enough to cover the seeds and moved to the greenhouse. Trays were sub-irrigated every two days to keep the soil moist, and plant height was measured for each plant 21 days after planting. The drc package (Ritz et al., 2015) in R (R-Core-Team, 2021) was used to fit 2-parameter log-logistic dose response curves for each population and herbicide.

Quantification of protoporphyrin accumulation

Plants from each tested population were grown to the 4-leaf stage. Plants were then sprayed with either oxyfluorfen (1122 g ai ha^{-1}; ALB2024; Albaugh LLC; Ankeny, IA, USA) or lactofen (841 g ai ha^{-1}; Cobra herbicide; Valent USA; San Ramon, CA, USA) using a moving overhead single-nozzle sprayer (DeVries Manufacturing, Hollandale, MN) calibrated to deliver 187 L ha^{-1}. Non-ionic surfactant was included in all herbicide mixtures at 0.25% v/v. Treated and untreated plants were stored in the dark immediately after herbicide treatment for 24 hours. At this point, the above ground tissue for four plants from each population and treatment was harvested and flash frozen in liquid nitrogen. Methods described by Takano et al. (2020) were used to extract and quantify protoporphyrin in each sample using LC/MS. Means of each population/treatment were compared using t-tests.

Tables

Table 4-1. KASP primer sequences used to detect premature stop codon mutations in Cadenza TILLING lines. The primer that binds to the wildtype allele has a HEX tail (green) and the primer that binds the mutant allele has a FAM tail (red). Thermocycler conditions of 95C for 15 minutes; 10 cycles of 95C for 20 seconds, 61C for 1 minute (-0.6C each round); 45 cycles of 95C for 20 seconds, X for 1 minute, read fluorescence at 34C for 1 minute should separate genotypes. Here, X represents the annealing temperature listed in the table.

Primer name	Sequence (5'-3')	Annealing Temperature
1D_UGP3_intron7F	CCTACCACTAAGATTCAGTCT	56
1D_UGP3_1626R_WT	GAAGGTCGGAGTCAACGGATTTTTGCGTGATCCAACTTCCTG	
1D_UGP3_1626R_MUT	GAAGGTGACCAAGTTCATGCTTTTGCGTGATCCAACTTCCTA	
1B_UGP3_intron5R	CAGTTATTCCCCCTTCCAAAC	56
1B_UGP3_1298F_WT	GAAGGTCGGAGTCAACGGATTGAAAAGGTGCAACTGTTCGTC	
1B_UGP3_1298F_MUT	GAAGGTGACCAAGTTCATGCTGAAAAGGTGCAACTGTTCGTT	
1A_UGP3_1168R	CTTCAGCATTTACTACAGGT	58.5
1A_UGP3_1080F_WT	GAAGGTCGGAGTCAACGGATTATCTGTGAAAGACTTGAATGG	
1A_UGP3_1080F_MUT	GAAGGTGACCAAGTTCATGCTATCTGTGAAAGACTTGAATGA	

Table 4-2. Concentrations of SQDG in young plant tissue of wildtype or *UGP3* knockout lines as quantified through LC/MS. The four values for each line represent biological replicates.

CRISPR		TILLING	
Sample	SQDG concentration (μg/ml)	Sample	SQDG concentration (μg/ml)
Cadenza0-1	14.16	Bobwhite-1	24.99
Cadenza0-2	17.98	Bobwhite-2	23.66
Cadenza0-3	19.88	Bobwhite-3	17.68
Cadenza0-4	21.02	Bobwhite-4	14.98
Mutant-1	0.1524	Mutant-1	0.849
Mutant-2	0.1839	Mutant-2	0.8261
Mutant-3	0.0994	Mutant-3	1.0854
Mutant-4	0.1752	Mutant-4	1.0129

Table 4-3. Parameter estimates for 2-parameter log-logistic dose response curves for wildtype (Bobwhite) and UGP3 knockout (CRISPR-M) populations in response to oxyfluorfen or lactofen. The equation fitted for each population was $y = \frac{100}{1+b(\log(x)-\log(ED50))}$ where y is predicted height as a percent of the untreated control 21 days after planting, x is the rate of herbicide used in a seed soak in mM, *ED50* is the rate of herbicide required to cause 50% reduction in height, and *b* is the slope of the curve at x=ED50.

Parameter	Estimate	Std. Error
Slope: Bobwhite (oxyfluorfen)	0.58	0.22
Slope: CRISPR-M (oxyfluorfen)	0.73	0.18
Slope: Bobwhite (lactofen)	1.95	0.41
Slope: CRISPR-M (lactofen)	2.07	0.86
ED50: Bobwhite (oxyfluorfen)	0.07	0.04
ED50: CRISPR-M (oxyfluorfen)	0.39	0.11
ED50: Bobwhite (lactofen)	0.51	0.08
ED50: CRISPR-M (lactofen)	0.82	0.10

Table 4-4. Sequences of sgRNA oligonucleotides. Tails with homology to type II restriction sites in the JD633 vector are indicated in green.

Name	Sequence (5’-3’)
417F	ACTTGCCAGGACCACGTGCTCGGCG
417R	AAACCGCCGAGCACGTGGTCCTGGC
919F	ACTTGCCGCATTGCTTCCTTACTG
919R	AAACCAGTAAGGAAGCAATGCGGC
1393F	ACTTGCATCTTGTGAGCGAAGACC
1393R	AAACGGTCTTCGCTCACAAGATGC

Table 4-5. Primer sequences for validation of JD633 vector construction. Thermocycler conditions of 95C for 5 minutes; 35 cycles of 95C for 30 seconds, 57C for 30 seconds, 72C for 1 minute; 72C for 7 minutes should produce an amplicon of ~800bp for non-digested JD633 and ~360bp for JD633 with a sgRNA sequence correctly inserted.

Name	Sequence (5’-3’)
TaU6promoter.F	TAGGAGGGAATCGAACTAGG
gRNAscaffold.R	CTTTTTCAAGTTGATAACGG

Table 4-6. Primer sequences used to amplify the edited locus of UGP3 homoeologs. The forward primers bind to sequence that is specific to each homoeolog and the reverse primer binds to a conserved region of exon 8. Lower case letters represent mismatches compared to the binding site. These mismatches were used to increase specificity or reduce secondary structure of each primer. Thermocycler conditions of 95C for 5 minutes; 35 cycles of 95C for 15 seconds, 60C for 15 seconds, 72C for 30 seconds; 72C for 5 minutes should produce an amplicon of ~300bp.

Primer name	Sequence (5’-3’)
1A_UGP3_intron_6F	TTGCACCTCCcGATTCTT
1B_UGP3_intron_6F	CGTCACAATAAGGTAAGAAAAAaGC
1D_UGP3_intron_6F	CGTCACAATAAGGTAAGAAAAAaGG
1ABD_UGP3_1537R	TTGCTGTAGGTTCTGATATGC

Figures

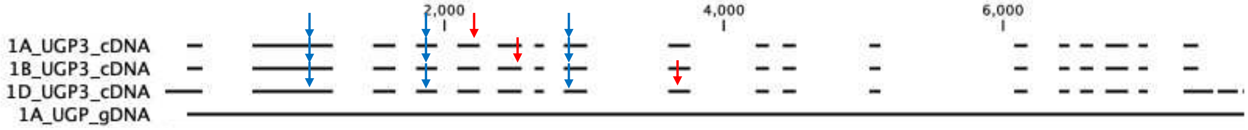
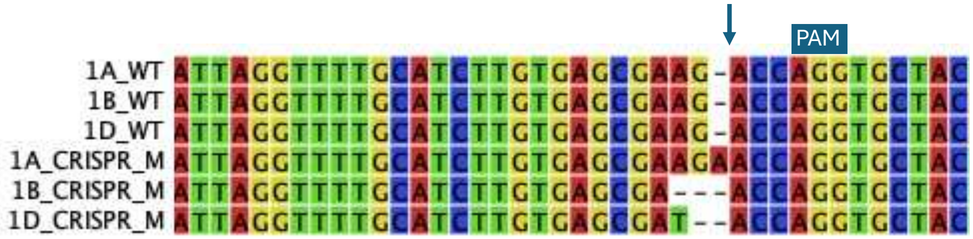


Figure 4-1. Schematic showing homology across *UGP3* homoeologs in wheat, gene structure, and sites of knockout mutations. Sites targeted by CRISPR-Cas9 editing system are indicated by blue arrows. Sites where premature stop codons were introduced in TILLING lines are indicated by red arrows.



```
1A_WT        ATTAGGTTTTGCATCTTGTGAGCGAAG-ACCAGGTGCTAC
1B_WT        ATTAGGTTTTGCATCTTGTGAGCGAAG-ACCAGGTGCTAC
1D_WT        ATTAGGTTTTGCATCTTGTGAGCGAAG-ACCAGGTGCTAC
1A_CRISPR_M  ATTAGGTTTTGCATCTTGTGAGCGAAGAACCAGGTGCTAC
1B_CRISPR_M  ATTAGGTTTTGCATCTTGTGAGCGA---ACCAGGTGCTAC
1D_CRISPR_M  ATTAGGTTTTGCATCTTGTGAGCGAT--ACCAGGTGCTAC
```

Figure 4-2. Sequence of *UGP3* homoeologs in wildtype and gene edited plants. The protospacer adjacent motif (PAM) is labeled with a blue box and the expected cut site for CRISPR-Cas9 is indicated with a blue arrow.

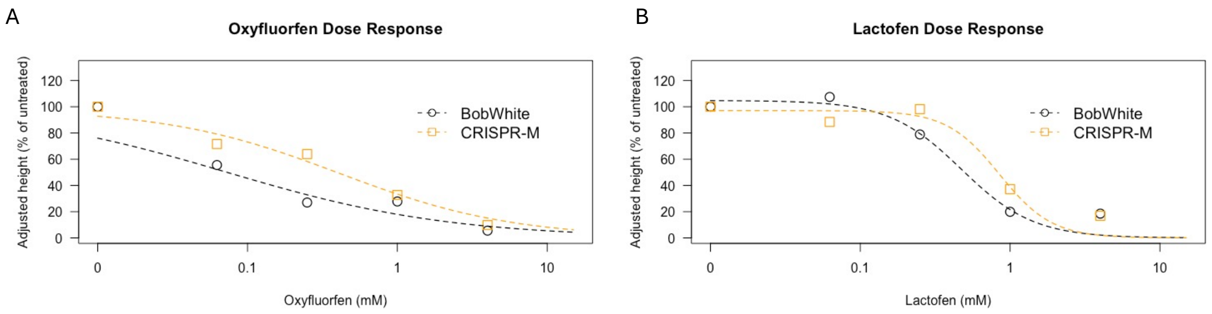


Figure 4-3. Dose response curves of wildtype (BobWhite) and *UGP3* knockout lines (CRISPR-M) to oxyfluorfen (A) or lactofen (B).

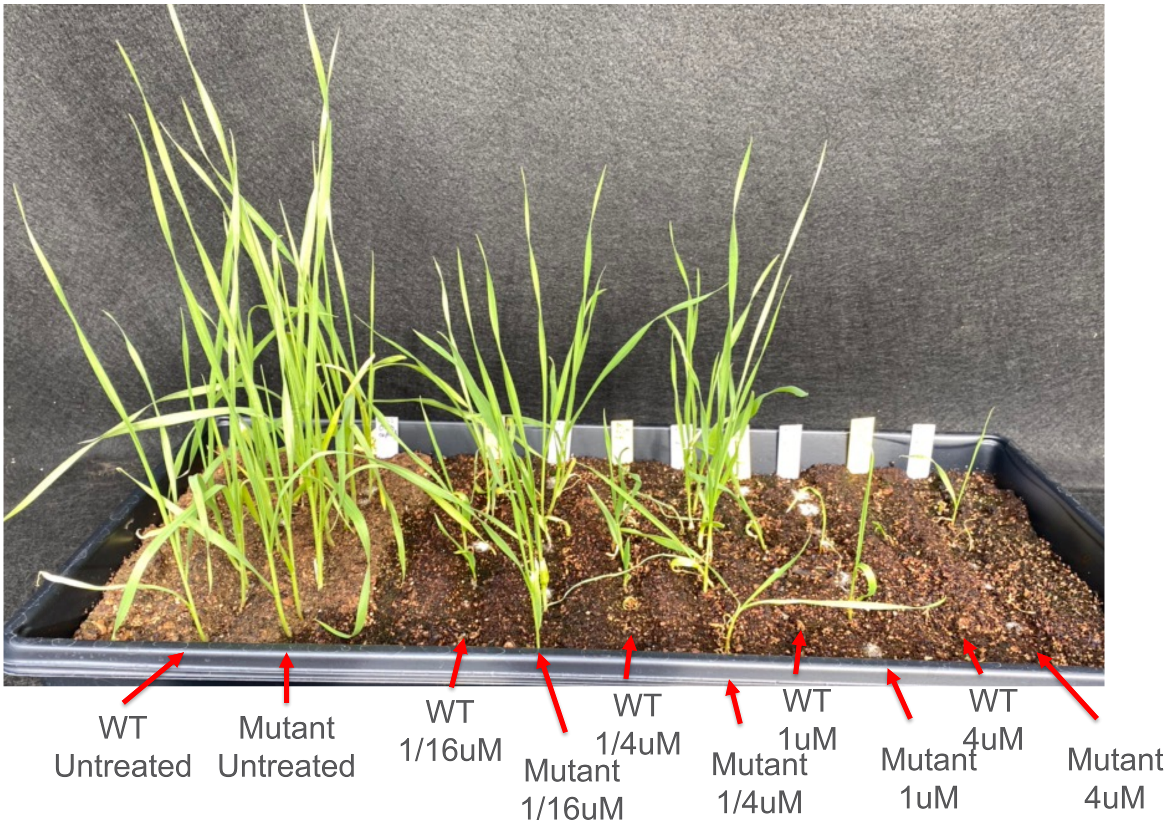


Figure 4-4. Seed soak dose response comparing Bobwhite (WT) and *UGP3*-edited (Mutant) wheat responses to oxyfluorfen. Photo taken 21 days after planting.

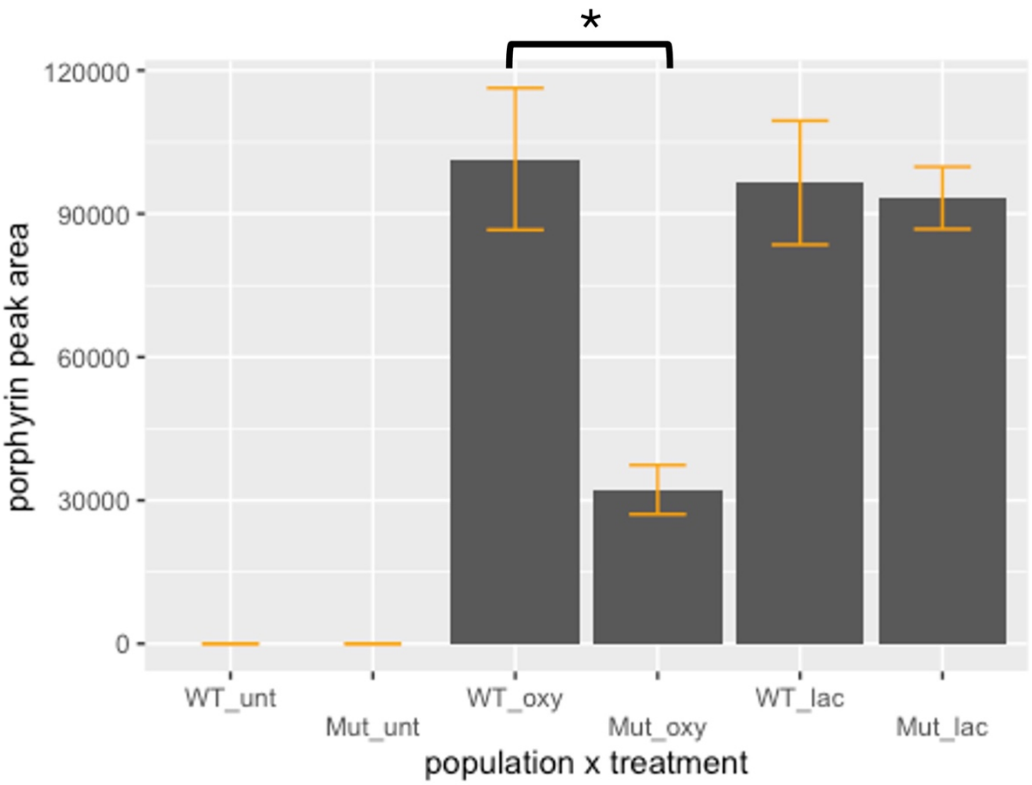


Figure 4-5. Concentration of protoporphyrin in young plant tissue of either wildtype (WT) or *UGP3* knockout lines (Mut) following application of either oxyfluorfen (oxy) or lactofen (lac) compared to an untreated control (unt). *=$p<0.05$.

References

Barker, A.L., Pawlak, J., Duke, S.O., Beffa, R., Tranel, P.J., Wuerffel, J. et al. (2023) Discovery, mode of action, resistance mechanisms, and plan of action for sustainable use of Group 14 herbicides. *Weed Science*, **71**, 173-188.

Bi, B., Wang, Q., Coleman, J.J., Porri, A., Peppers, J.M., Patel, J.D. et al. (2020) A novel mutation A212T in chloroplast Protoporphyrinogen oxidase (PPO1) confers resistance to PPO inhibitor Oxadiazon in Eleusine indica. *Pest Manag Sci*, **76**, 1786-1794.

Buchholzer, M. & Frommer, W.B. (2023) An increasing number of countries regulate genome editing in crops. *New Phytologist*, **237**, 12-15.

Busi, R., Vila-Aiub, M.M., Beckie, H.J., Gaines, T.A., Goggin, D.E., Kaundun, S.S. et al. (2013) Herbicide-resistant weeds: from research and knowledge to future needs. *Evolutionary Applications*, **6**, 1218-21.

Dayan, F.E., Weete, J.D., Duke, S.O. & Hancock, H.G. (1997) Soybean (Glycine max) cultivar differences in response to sulfentrazone. *Weed Science*, **45**, 634-641.

Debernardi, J.M., Tricoli, D.M., Ercoli, M.F., Hayta, S., Ronald, P., Palatnik, J.F. et al. (2020) A GRF-GIF chimeric protein improves the regeneration efficiency of transgenic plants. *Nat Biotechnol*, **38**, 1274-1279.

Duke, S.O. & Dayan, F.E. (2022) The search for new herbicide mechanisms of action: Is there a 'holy grail'? *Pest Manag Sci*, **78**, 1303-1313.

Hao, G.F., Zuo, Y., Yang, S.G. & Yang, G.F. (2011) Protoporphyrinogen oxidase inhibitor: an ideal target for herbicide discovery. *Chimia (Aarau)*, **65**, 961-9.

Heap, I., The International Herbicide-Resistant Weed Database. www.weedscience.org [accessed 20 June 2024].

Johnson, K., Cao Chu, U., Anthony, G., Wu, E., Che, P. & Jones, T.J. (2023) Rapid and highly efficient morphogenic gene-mediated hexaploid wheat transformation. *Front Plant Sci*, **14**, 1151762.

Kerr, D.R., Concepcion, J.C.T., Strom, S.A. & Riechers, D.E. (2023) Quantifying resistance to very-long-chain fatty acid-inhibiting herbicides in Amaranthus tuberculatus using a soilless assay. *PLOS ONE*, **18**, e0295927.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissold, L. et al. (2017) Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences*, **114**, E913-E921.

Larue, C.T., Ream, J.E., Zhou, X., Moshiri, F., Howe, A., Goley, M. et al. (2020) Microbial HemG-type protoporphyrinogen IX oxidase enzymes for biotechnology applications in plant herbicide tolerance traits. *Pest Manag Sci*, **76**, 1031-1038.

Lermontova, I. & Grimm, B. (2000) Overexpression of plastidic protoporphyrinogen IX oxidase leads to resistance to the diphenyl-ether herbicide acifluorfen. *Plant Physiol*, **122**, 75-84.

Liu, H., Ding, Y., Zhou, Y., Jin, W., Xie, K. & Chen, L.-L. (2017) CRISPR-P 2.0: An Improved CRISPR-Cas9 Tool for Genome Editing in Plants. *Molecular Plant*, **10**, 530-532.

Nakajima, Y., Umena, Y., Nagao, R., Endo, K., Kobayashi, K., Akita, F. et al. (2018) Thylakoid membrane lipid sulfoquinovosyl-diacylglycerol (SQDG) is required for full functioning of photosystem II in Thermosynechococcus elongatus. *J Biol Chem*, **293**, 14786-14797.

Newhouse, K.E., Smith, W.A., Starrett, M.A., Schaefer, T.J. & Singh, B.K. (1992) Tolerance to imidazolinone herbicides in wheat. *Plant Physiol*, **100**, 882-6.

Nie, H., Harre, N.T. & Young, B.G. (2023) A New V361A Mutation in Amaranthus palmeri PPX2 Associated with PPO-Inhibiting Herbicide Resistance. *Plants (Basel)*, **12**, Ostlie, M., Haley, S.D., Anderson, V., Shaner, D., Manmathan, H., Beil, C. et al. (2015) Development and characterization of mutant winter wheat (Triticum aestivum L.) accessions resistant to the herbicide quizalofop. *Theor Appl Genet*, **128**, 343-51.

Patzoldt, W.L., Hager, A.G., McCormick, J.S. & Tranel, P.J. (2006) A codon deletion confers resistance to herbicides inhibiting protoporphyrinogen oxidase. *Proceedings of the National Academy of Sciences*, **103**, 12329-12334.

R-Core-Team, R: A Language and Environment for Statistical Computing. https://www.R-project.org/ [accessed

Rangani, G., Salas-Perez, R.A., Aponte, R.A., Knapp, M., Craig, I.R., Mietzner, T. et al. (2019) A Novel Single-Site Mutation in the Catalytic Domain of Protoporphyrinogen Oxidase IX (PPO) Confers Resistance to PPO-Inhibiting Herbicides. *Front Plant Sci*, **10**, 568.

Ritz, C., Baty, F., Streibig, J.C. & Gerhard, D. (2015) Dose-response analysis using R. *PLOS ONE*, **10**, e0146021.

Rousonelos, S.L., Lee, R.M., Moreira, M.S., VanGessel, M.J. & Tranel, P.J. (2012) Characterization of a Common Ragweed (Ambrosia artemisiifolia) Population Resistant to ALS- and PPO-Inhibiting Herbicides. *Weed Science*, **60**, 335-344.

Soni, N., Westra, E.P., Allegretta, G., Araujo, A.L.S., de Pinho, C.F., Morran, S. et al. (2022) Survey of ACCase and ALS resistance in winter annual grasses identifies target-site and nontarget-site imazamox resistance in Secale cereale. *Pest Manag Sci*, **78**, 5080-5089.

Tadesse, W., Sanchez-Garcia, M., Assefa, S.G., Amri, A., Bishaw, Z., Ogbonnaya, F.C. et al. (2019) Genetic Gains in Wheat Breeding and Its Role in Feeding the World. *Crop Breeding, Genetics and Genomics*, **1**, e190005.

Takano, H., Beffa, R., Preston, C., Westra, P. & Dayan, F. (2020) Glufosinate enhances the activity of protoporphyrinogen oxidase inhibitors. *Weed Science*, **68**, 324-332.

Vencill, W.K., Nichols, R.L., Webster, T.M., Soteres, J.K., Mallory-Smith, C., Burgos, N.R. et al. (2012) Herbicide Resistance: Toward an Understanding of Resistance Development and the Impact of Herbicide-Resistant Crops. *Weed Science*, **60**, 2-30.

Wang, K., Shi, L., Liang, X., Zhao, P., Wang, W., Liu, J. et al. (2022) The gene TaWOX5 overcomes genotype dependency in wheat genetic transformation. *Nat Plants*, **8**, 110-117.

Weigel, D. & Glazebrook, J. (2002) How to transform *arabidopsis*. In *Arabidopsis* eds). Cold Spring, NY, USA: Cold Spring Harbor Laboratory Press, pp.

Yu, B., Xu, C. & Benning, C. (2002) *Arabidopsis* disrupted in *SQD2* encoding sulfolipid synthase is impaired in phosphate-limited growth. *Proceedings of the National Academy of Sciences*, **99**, 5732-5737.

Zhou, H., Arrowsmith, J.W., Fromm, M.E., Hironaka, C.M., Taylor, M.L., Rodriguez, D. et al. (1995) Glyphosate-tolerant CP4 and GOX genes as a selectable marker in wheat transformation. *Plant Cell Rep*, **15**, 159-63.

Zhu, T., Wang, L., Rimbert, H., Rodriguez, J.C., Deal, K.R., De Oliveira, R. et al. (2021) Optical maps refine the bread wheat Triticum aestivum cv. Chinese Spring genome assembly. *Plant J*, **107**, 303-314.